

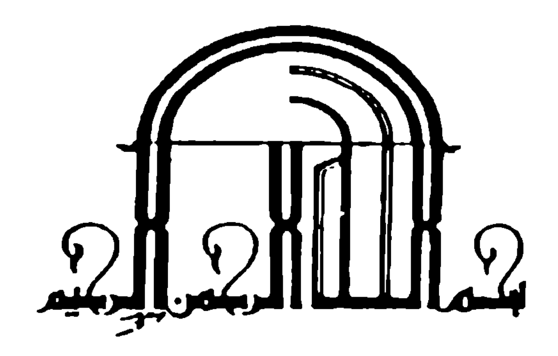
مطبوعات
مكتبة الملك فهد الوطنية
السلسلة الثانية
(٥١)

محركات بحث الصور الثابتة على الإنترنت

دراسة تحليلية

سيد ربيع سيد إبراهيم

الرياض
١٤٢٨هـ/ ٢٠٠٧ م



محركات بحث الصور الثابتة

على الإنترنت

مطبوعات

مكتبة الملك فهد الوطنية

السلسلة الثانية

(٥١)

تعنى هذه السلسلة بنشر الدراسات والبحوث
في إطار علم المكتبات والمعلومات بشكل عام

محركات بحث الصور الثابتة على الإنترنت

دراسة تحليلية

سيد ربيع سيد إبراهيم

معيد بقسم المكتبات والوثائق

مكتبة الملك فهد الوطنية

الرياض ١٤٢٨ هـ - ٢٠٠٧ م

© مكتبة الملك فهد الوطنية، ١٤٢٨ هـ

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

إبراهيم، سيد ربيع سيد
محركات بحث الصور الثابتة على الإنترنت / سيد ربيع
سيد إبراهيم - الرياض، ١٤٢٨ هـ
٣٤٤ ص : ٢٤ سم (الثانية ؛ ٥١)
ردمك : ٨-٢٨٨-٠٠-٩٩٦٠

١ - محركات البحث ٢- الانترنت أ. العنوان ب. السلسلة
ديوي ٠٠٤،٦٧٨ ١٤٢٨/٢٣٨٥

رقم الإيداع : ١٤٢٨/٢٣٨٥
ردمك : ٨-٢٨٨-٠٠-٩٩٦٠

ص . ب : ٧٥٧٢
الرياض : ١١٤٧٢ - المملكة العربية السعودية
هـاتف : ٤٦٢٤٨٨٨
فاكس : ٤٦٤٥٣٤١

المحتويات

الموضوع	الصفحة

الباب الأول : الإطار المنهجي للدراسة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة



قائمة الجداول

الجداول | **الصفحة**



مقدمـــة الدراســة

المقدمــة

في ظل عالم تتزاحم فيه أشكال أوعية المعلومات ما بين مخطوط ومطبوع، ومواد سمعية وبصرية، إلى وسائط متعددة وشبكات معلومات، تمثل الصورة المرئية الثابتة والمتحركة مصدرًا مهمًا من مصادر المعلومات الأولية بجوار المادة النصية والمادة الصوتية.

وتعد الصور ذات أهمية كبرى سواء للأغراض التعليمية أو البحثية أو أغراض عرض المعلومات لفئة خاصة من المستفيدين مثل المعاقين سمعيا، كما تخدم الصور الصحفيين في التعبير عن الأحداث. ويحتاج إليها أيضًا في المجالات العلمية مثل: الطب والكيمياء والفلك والنبات والحيوان والطبيعة وغيرها. ونظرًا لما تتضمنه الصورة من خصائص بشكلها المرئي، فهي أكثر تأثيرًا في داخل المتلقي حيث تغني صورة لحدث ما عن وصفه في ألف كلمة. وهو الأمر الذي جعل الصور تمثل أداة تعليم أساسية في رياض الأطفال، كما أن مناهج التعليم المهني تعتمد أيضًا على الكتب والموسوعات التي تجعل مادتها المصورة هي الأصل والنص مساعدًا لها، ومن ذلك كتب الطهي والتطريز والصيانة... وغيرها.

وتعد الصور أكثر أشكال أوعية المعلومات ندرة في مؤسسات المعلومات. واتبع هذه الندرة قلة الاهتمام بطرح مواصفات اختزان واسترجاع ووصف الصور بما يتناسب مع

خواص الصورة كوعاء له أهميته في عرض المعلومات المرئية، ولا يوجد من قواعد الفهرسة الأنجلو أمريكية AACR2 ما يصف التسجيلات المرئية؛ إلا ما اختص بالفهرس البطاقي دون الفهرس الآلي. كما غاب عن التقعيد وصف أساليب استرجاع الصور في قواعد البيانات الآلية. وهو ما لا يتناسب مع العدد الذي توجد به قواعد بيانات الصور العامة والعلمية المتخصصة.

وتختلف الصورة عن النص في الشكل الورقي من حيث أساليب التنظيم عن النص في مناهج التكشيف واستخدام الواصفات، حيث يتم تحليل الصورة وفق موضوعها أو السياق الذي وردت فيه. ويتحدد في ذلك مستويات التكشيف المختلفة سواء العناصر الأساسية البارزة في الصورة أو الخلفية أو الصورة الذهنية التي تتشكل في ذهن المشاهد خارج إطار العناصر الموضوعية لها. وهي كلها مشكلات تمثل عائقا كبيرًا في استرجاع الصور بخلاف النص الذي تحدده الكلمات. أما في الاسترجاع الآلي فتتضافر عناصر أخرى تجعل أمر التكشيف على مستوى النص أو الصورة أكثر صعوبة. فالاسترجاع الآلي يعتمد على عدد من الآليات مثل المنطق البوليني والبتر الضمني يتحدد وفق خصائص اللغة المستخدمة.

يضاف إلى ما سبق أن الاسترجاع الآلي للنص يعتمد على الموضوع فقط وهو ما تعكسه مفردات اللغة المستخدمة. أما في الصور فإن الاسترجاع يتم وفق الخصائص المادية المصاحبة للصورة أيضا والتي تتمثل في اللون والشكل والتركيب، وهو ما يتطلب تحليل الصور بناء على الخصائص المادية لها بالإضافة إلى الموضوعية وصياغة تسجيلة معيارية تجمع بين واصفات التحليل الموضوعي وواصفات التحليل المادي للصورة. هذا إلى جانب الآليات التي يعمل بها محرك بحث الويب تتمثل في برامج الاختيار والتنظيم والبحث تناسب أكثر واقع تنظيم واستدعاء المصادر النصية التي ينتهي امتدادها على الويب إلى HTM و HTML، مما يعكس أن التعامل مع ملفات الصور الرقمية إنما ينبني فقط على النص المصاحب وهو ما يتفاوت في أحيان كثيرة بين الكلمات المكونة لاسم ملف الصور إلى بضع جمل واصفة لتحليل الخصائص المادية والموضوعية إلى الصورة، وهو المستوى الأكثر غيابًا عن بناء ملفات الصور.

أولًا : مشكلة الدراسة

تذخر الإنترنت بالكثير من أنواع ملفات المعلومات الصورة. وأهم هذه الأنواع هو ملف الصور الرقمية، وتنبع مشكلة الدراسة من الكم الهائل الذي تتواجد به هذه الملفات على الإنترنت حتى وصلت إلى ملايين الصور الرقمية الثابتة على الإنترنت. وتواجه هذه الدراسة جوانب عدة تتعلق بمشكلة تنظيم واسترجاع ملفات الصور الرقمية على الإنترنت وأول هذه الجوانب هو عدم وجود معيار موحد مستخدم لتنظيم ومعالجة الصور الرقمية المعالجة الفنية على الإنترنت ؛ حيث يستخدم لوصف الصور الرقمية الكثير من عناصر وحقول الوصف الفني المتوافقة مع طبيعة الموقع أو الهيئة التي تستخدم مجموعات الصور الرقمية، وهذه الحقول لا تكفي للدرجة التي تحقق أعلى مستوى للاسترجاع من الصور الرقمية إلى جانب انخفاض مستوى التحقيق في نتائج الصور الرقمية المسترجعة على الإنترنت.

أما الجانب الثاني محركات البحث المتخصصة في بحث الصور الرقمية إلى أسلوب يتناسب مع إضافة وتنظيم وبحث ملفات الصور الرقمية ؛ حيث تتعامل محركات البحث مع هذه الملفات كغيرها من الملفات النصية معتمدة في تنظيم واسترجاع الصور الرقمية على النص المرافق للصورة دون المحتوى الموضوعي أو حقول المعالجة الفنية للصور الرقمية كوسيط لحمل المعلومات له خصائصه المختلفة كلية عن وسيط النص.

ويتعلق الجانب الثالث من مشكلة الدراسة بخصائص الصورة في حمل المعلومات كعناصر الوصف والتحليل الموضوعي الخاصة بها بما تحمله الصورة من أبعاد ثلاثة تختلف في عرض الموضوع وأعمقها هو البعد الثالث في محتوى الصورة الذي يرتبط بالمعنى المراد إيصاله من خلال عناصر الصورة المرئية، فضلا عن جوانب حقوق النشر وترشيح مجموعات الصور الرقمية التي تتميز به دون غيرها من وسائط المعلومات النصية على الإنترنت.

أما الجانب الرابع فيتعلق بواقع معالجة ملفات الصور الرقمية بشكل خاص في البيئة الرقمية العربية التي تفتقر إلى وجود معايير موحدة لوصف ومعالجة الصور الرقمية الثابتة، خاصة ووسائط حمل المعلومات المصورة عامة.

إن تكامل الجوانب السابقة لمشكلة الدراسة يوضح أن مشكلة الدراسة تتركز في البحث عن إيجاد معيار موحد يحمل حقول الوصف الملائمة للمعالجة الفنية للصور الرقمية على الويب، مع وضع مواصفات تفيد محركات بحث الصور في العمل بكفاية أقدر على تحقيق نسب الاسترجاع والتحقيق المرجوة من قبل المستفيدين، إلى جانب دراسة خصائص اللغة العربية للتعامل مع ملفات الصور الرقمية بما يفيد الباحث العربي؛ إضافة إلى إمكانية تكوين مجموعات الصور الرقمية العربية التي تتوافق في خصائص تنظيمها وبحثها مع خصائص اللغة العربية.

ثانيًا: تساؤلات الدراسة

يمكن صياغة وتحديد مشكلة الدراسة السابقة في الأسئلة التالية:-

١. ما التطور التاريخي لتناول الصور الرقمية في أدبيات الإنتاج الفكري وواقع المعالجة الفنية داخل مؤسسات المعلومات؟
٢. ما ملامح خصائص ملفات الصورة الرقمية مقارنة مع خصائص الصورة الورقية؟
٣. ما خصائص الصورة وطبيعتها كوسيط معلومات مرئي مقارنة بالنص؟
٤. ما أهم ملامح المعايير العالمية لتكشيف واسترجاع الصور الرقمية الثابتة؟
٥. كيف يمكن دعم آليات عمل محركات البحث المتخصصة بما يوافق بنية ملفات الصور الرقمية على الإنترنت؟
٦. ما مدى تأثير مرشحات محركات بحث الصور على نتائج البحث؟
٧. ما ملامح عمل مرشحات الصور الرقمية في البيئة العربية الإسلامية؟
٨. ما خصائص اللغة العربية في تكشيف الصور الثابتة واسترجاعها؟
٩. ما مدى كفاية محركات الصور الثابتة المعربة للتعامل مع خصائص اللغة العربية؟
١٠. ما مواصفات عمل محركات بحث الصور الثابتة التي تدعم اللغة العربية؟

ثالثًا: أهمية الدراسة ومبرراتها

تأتى أهمية الموضوع ومبررات اختياره للأسباب التالية:-

١. الاتجاه القوي داخل المكتبات ومرافق المعلومات نحو التعامل مع البيئة الإلكترونية وأوعية المعلومات الإلكترونية. ويتزايد هذا الاتجاه مع الأوعية الأكثر عرضة للتلف مع كثرة تداولها مثل الصور والخرائط والمخطوطات والكتب النادرة.

٢. تزايد أهمية الصور في كافة الميادين البحثية والتعليمية والإعلامية مع تقدم التقنيات المصاحبة لها؛ حيث أصبحت هناك صور ميكروسكوبية، وصور أشعة، وصور بالأقمار الصناعية مع تعدد أشكال كل منها، بالإضافة إلى تراجع النص أمام الصور بسبب التضخم الهائل في الإنتاج الفكري وقلة وقت الباحث والقارئ.

٣. تصاعد أهمية محركات البحث في ظل التزايد المطرد لاستخدام الإنترنت وتزايد المادة المتاحة من خلالها، وتعدد وتباين الدراسات التقييمية لمحركات البحث من منظور كفاية لاسترجاع النص، دون التطرق لكفاية استرجاع المادة المصورة أو الصوتية.

٤. ندرة الدراسات التي اتجهت نحو التقعيد والتوصيف للمعالجة الموضوعية للصور وخاصة الصور الرقمية، حيث اتجهت معظم الدراسات إلى التركيز على الاسترجاع على أساس المعالجة المادية، وانحصرت المعالجة الموضوعية في دراسة أنماط أسئلة وطلبات المستفيدين.

٥. عدم وجود مواصفات معيارية تخدم تكشيف الصور لأغراض البحث باللغة العربية على كلا المستويين الورقي والرقمي. كما أن محركات البحث العربية لا تدعم البحث في الصور أصلا. أما المحركات غير العربية فإنها تتعامل مع استرجاع الصور بالنص العربي من خلال آليات تعريب تفتقر إلى الدقة والكفاية والمعيارية.

رابعًا : أهداف الدراسة

تحاول الدراسة من خلال التحليل والتقييم لمحركات بحث الصور الوصول إلى تحقيق الأهداف التالية:-

١. التعرف إلى الإنتاج الفكري المنشور في مجال استرجاع الصور الرقمية الثابتة.
٢. التعرف إلى خصائص الصورة الرقمية مقارنة مع خصائص الصورة الورقية.
٣. التعرف إلى خصائص الصورة وطبيعتها كوعاء معلومات مرئي مقارنة بالنص.
٤. تحديد معايير واصفات تكشيف واسترجاع الصور الرقمية الثابتة.
٥. حصر وتقييم آليات عمل محركات البحث المتخصصة في الصور الثابتة.
٦. قياس مدى كفاية محركات بحث الصور ذات مرشحات الويب أو بدونها على نتائج البحث.
٧. دراسة خصائص اللغة العربية في تكشيف الصور الثابتة واسترجاعها.
٨. قياس مدى كفاية محركات الصور الثابتة المعربة للتعامل مع خصائص اللغة العربية.
٩. وضع مواصفات محركات بحث الصور الثابتة التي تدعم اللغة العربية.

خامسًا : تعريفات إجرائية

- **محرك بحث:**

هو عبارة عن برنامج متكامل لتنظيم وبحث مصادر الويب، من خلال العمل بثلاثة برامج فرعية هي الزاحف لاقتناء صفحات الويب، والمكشف لتنظيم وتكشيف صفحات الويب، وآليات البحث لدعم استرجاع صفحات الويب بحسب الحاجة الموضوعية للمستفيدين.

- **المادة المصورة:**

يحوي مصطلح المادة المصورة كل مصادر معلومات الويب المرئية التي تعتمد على وسيط الرؤية لنقل المعلومات إلى المستفيد. ويستخدم هذا المصطلح داخل الدراسة ليشمل المواد المرئية المتحركة وغير المتحركة.

- **الصور الثابتة:**

هي مصدر المعلومات المرئي الثابت المكون من عناصر مرئية أو شكلية متكاملة. سواء كان شكلاً تصويريًا أو صورة فوتوغرافية أو رسومًا جغرافية...الخ. ويقصد به في الدراسة الصور الفوتوغرافية الرقمية.

سادسًا : عينة الدراسة

اهتمت الدراسة بتتبع خصائص مجتمع البحث العام المتمثل في مجتمع محركات البحث على الويب ومجتمع الصور الرقمية على الويب، ثم اختيار عينة ممثلة لهذا المجتمع، وكانت عينة الدراسة كالتالي:

تنقسم العينة المستخدمة في الدراسة إلى الفئات التالية:

- **عينة محركات البحث:**

اعتمدت الدراسة على عينة قصدية من محركات البحث التي تبحث في الصور الثابتة سواء تلك المتخصصة في حفظ واسترجاع الصور الرقمية أو التي تجعل من بحث الصور اختيارًا إضافيًا في بحث مصادر الويب. وتم اختيار العينة بالاستعانة بالدراسات الإحصائية المتاحة على الإنترنت لقياس كفاية محركات البحث، وتم اختيار محركات البحث التي تحقق نسبة أعلى من ٥٠٪ في درجات التقييم[١][٢].

وقد انقسمت أنواع محركات البحث في عينة الدراسة إلى نوعين هما:-

محركات بحث الويب

ويختص هذا النوع بالتركيز على بحث الصور الرقمية الثابتة دون غيرها من مصادر الويب.

---

(١) http://www.search-engine-index.co.uk/

(٢) http://www.sarchenginespy.co.uk/update/pages/national/htm

اسم محرك البحث	مسار محرك البحث
Alltheweb "Fast"	http://multimedia.alltheweb.com
Altavista	http://www.altavista.com/image/default?qbmode=
Ditto	http://www.ditto.com
Google	http://images.google.com
Picsearch	http://www.picsearch.com
Visoo	http://www.visoo.com/

محركات بحث محددة

ويختص هذا النوع بالمحركات التي تمثل قواعد معلومات صور ثابتة.

اسم محرك البحث	مسار محرك البحث
Corbis	http://www.corbis.com
Freefoto	http://www.freefoto.com

■ نماذج الكلمات البحثية

اعتمدت الدراسة على العينة التالية من كلمات البحث لتحليل خصائص محركات البحث:-

المصطلح	مكان الاستخدام	عنصر الدراسة
Egypt and pyramids	آليات البحث	معاملات المنطق البوليني And, or, not
Egypt or pyramids		
Egypt not pyramids		
EARTH	آليات البحث	حساسية الحالة أو الحروف
Earth		
plane + sky	آليات البحث	بدائل المنطق البوليني
plane - sky		
"sea beach"	آليات البحث	التطابق
Egypt > 1995	آليات البحث	التغطية الزمنية
Egypt < 1995		
Egypt near3 pyramids	آليات البحث	البحث بحالة الكلمات
Egypt adj pyramids		
Egy?t	آليات البحث	البحث الحر
Egypt*		
Sky	مستويات التكشيف	الخلفية والبعد الثالث في الصورة
Planet sky		

المصطلحات العربية:

المصطلح	مكان الاستخدام	عنصر الدراسة
الأهرام مصر	آليات البحث	معاملات المنطق البوليني And, or, not
الأهرام		
مصر		
الشاطئ البحر	آليات البحث	التطابق
القمر، قمر	آليات البحث	اللواصق
سيارة	آليات البحث	الشوشرة (كفاية الاسترجاع)
طائرة	آليات البحث	الشوشرة (كفاية الاسترجاع)
قطر	آليات البحث	الشوشرة (كفاية الاسترجاع)

- **أدوات الدراسة التجريبية**

اعتمد الباحث في التعامل مع قياس كفاية محركات بحث الصور في خصائص البحث والاسترجاع على مجموعة ثابتة من التجهيزات، واستخدمت هذه التجهيزات بشكل ثابت أثناء التعامل مع كل محركات عينة الدراسة وهي كالتالي :-

١. جهاز حاسب آلي بإمكانات هي :

- معالج نوع p4 "MGH 1200"
- لوحة أساسية Main Board FIC
- كارت kb 56 fax modem ــــ شاشة 15" viewer

٢. خط هاتف بخاصية dial up connection

وقد عمل الباحث على حساب النتائج باستخدام الأرقام المسترجعة للبحث، تقدير الاختلافات بين أعداد النتائج المسترجعة من قواعد بيانات محركات البحث.

سابعًا: منهج الدراسة وأدواتها

اعتمدت الدراسة على استخدام عدد من المناهج منها :

- المنهج الوصفي الذي يعتمد على تتبع الإنتاج الفكري وتحليل نتائج الدراسات الخاصة في مجال تكشيف الصور؟

- المنهج التجريبي في استخدام نماذج مختلفة من الواصفات مع محركات الصور وجمع نتائج الاسترجاع وتحليلها، ثم قياس مدى فاعلية هذه النتائج مع أساليب الاسترجاع الآلي المطروحة للصور الرقمية الثابتة، وخصائص اللغة العربية ووضع معايير ومواصفات لمحركات بحث تدعم اللغة العربية.
- المنهج المقارن الذي يعتمد على المقارنة بين كفاية محركات البحث في الاستدعاء والاسترجاع.

أما أدوات جمع البيانات فتتمثل في

- **أداة البحث الوثائقي** لجمع الجانب النظري من الدراسة التي تتعلق ببناء الويب وعمل محركات البحث في جمع المصادر، إلى ما يتعلق بملفات الصور الرقمية وخصائصها.
- **أداة المعايشة مع الإنترنت** من خلال مشاهدة ومراقبة أساليب البحث والاسترجاع في محركات البحث وتقييم عمل هذه الأساليب والنتائج التي ستخرج منها مصطلحات البحث .

ثامنًا : حدود الدراسة

- الحدود الموضوعية:

تقتصر الدراسة على موضوع محركات البحث على الويب، المحددة التي تبحث في الصور ذات الصبغة العامة على الإنترنت دون التعرض لمحركات البحث الإعلامية الخاصة بوكالات الأنباء أو بمكتبات الصور الرقمية العلمية والأدلة ذات التخصص الدقيق التي تبحث في الصور العلمية أو التاريخية أو الجغرافية.

- الحدود النوعية:

تقتصر الدراسة على الصور الثابتة، ولا يدخل في ذلك أي من الصور المتحركة المتمثلة في لقطات الفيديو أو الرسوم المتحركة.

- الحدود الشكلية:

تقتصر الدراسة على الصور الرقمية الفوتوغرافية بشكل خاص، ولا يدخل في ذلك أي من أنواع المادة المصورة الأخرى مثل الخرائط، أو الرسوم الكاريكاتيرية، أو اللوحات الفنية أو غيرها من الأشكال.

- الحدود اللغوية:

تقتصر الدراسة على محركات بحث الصور التي تدعم اللغتين الإنجليزية والعربية فقط دون التعرض لمحركات البحث التي تبحث في اللغات الأخرى.

- الحدود الزمنية:

تقتصر الدراسة على محركات البحث المتاحة على الويب حتى نهاية شهر أبريل ٢٠٠٥م

تاسعًا : الدراسات المثيلة

أ - دراسات سابقة للصور الثابتة:-

عمل الباحث على مسح مصادر المعلومات التقليدية والرقمية التي تتعلق بتنظيم وبحث المادة المصورة الرقمية في بيئة محركات البحث. وقد أثرت الخلفية التاريخية للإنتاج الفكري الذي بدأ فقط في النصف الثاني من عقد التسعينات على قلة توافر المصادر التقليدية في هذا الموضوع. وكان السواد الأعظم من مصادر المعلومات في الشكل الرقمي "مصادر الويب". واقتصرت هذه المصادر على مقالات الدوريات على الويب والأوراق البحثية المقدمة في مؤتمرات مناقشة تنظيم مصادر الويب، وأيضا المعلومات الإلكترونية في هيئات تنظيم الويب والصور الرقمية، والدراسات الأكاديمية النادرة التي تتعلق بتنظيم معلومات الصور.

كما تبين ندرة بعض قواعد البيانات المتخصصة في ذلك مثل: Dissertation Abstracts International UMI. إلا أن عمليات بحث هذه القواعد قد أشارت إلى ندرة الاهتمام بتنظيم الصورة كوعاء للمعلومات، في حين اهتم معظمها بما يتعلق بالصور الرقمية والتخزين على الحاسبات الآلية وما يتعلق بها من تقنيات مثل Content Based Image Retrieval؛ حيث استخدام الألوان والشكل لحفظ وتنظيم الصورة الرقمية بعيدًا عن مفاهيم التكشيف أو المعالجة الفنية للمعلومات. ومن الدراسات ذات العلاقة في الإنتاج الفكري الأجنبي ما يلي:-

الدراسة الأولى:

Jorgensen Corinne. Image Attributes: An Investigation, Syracuse University, 1995.- 323p.- Ph.D.

وتتناول هذه الدراسة أدبيات علم النفس وعلم الجمال وعلم تكشيف الصور،

وعلم النفس المعرفي كعلوم أساسية تساهم في وصف نماذج الاستخدام المختلفة للمادة المصورة. وقد تم تحليل عناصر٤٧ صورة بشكل إحصائي من خلال تفسير نتائج المشاركين في التعامل مع الصور من منظور الوصف، والتصنيف، والاسترجاع. كما تم دراسة اللون والمحتوى والعناصر الانطباعية بهدف التعرف إلى تكشيف الصور لأغراض تكشيف القصص الخيالية.

الدراسة الثانية:

Ahmed Imran. A Hierarchical Decomposition Approach for Image Indexing, Advisor Grosky William, Wayne ÚState University, 1997.- 104p.- Ph.D.

تطرح هذه الرسالة تقنية تمثيل رمزي جديد للصور مع طريقة تكشيف لاسترجاع الصور على أساس التشابه الفراغي فيها. وتحلل هذه الطريقة الصورة في ترتيب فراغي للسمات المختلفة فيها بشكل تكراري، والاحتفاظ في العلاقات الفراغية بين عناصر الصور المتنوعة. ومبدأ الطريقة هو عمل تحليل هرمي يساعد في تحقيق قياس التشابه في عناصر الصورة مع تراكم يسمح بإيجاد نسبة تطابق إلى مستويات متباينة من التفاصيل نزولا من العامة إلى الدقيقة. وفي كل مرحلة يتم فيها التطابق تعطي الصورة اسم ملف يمكن به حذفها من المقارنة في حالة عدم تطابقها.. بينما يتم ترتيب الصور المتبقية والمطابقة تنازليا من الأكثر مطابقة إلى الأقل مطابقة وفق عناصر التشابه بما يسمح بتصفح الصور في قاعدة البيانات والاسترجاع من قبل المستفيد.

الدراسة الثالثة:

Huan Jing. Color Spatial Image Indexing and Applications, Advisor Zabih Ramin.- Cornell University, 1998, 156p.- Ph.D.

تستخدم هذه الدراسة تقنية جديدة لتكشيف الصور على أساس العلاقة الفراغية للألوان وللصورة أطلق عليها مسمى correlogram وتتناول التقنية المشكلات المتنوعة التي تظهر في استرجاع الصورة على أساس المحتوى وتصفح لقطات الفيديو. وقد تم تطبيق التقنية على قاعدة بيانات تحتوي على ٢٠٠ ألف صورة وتعتمد هذه التقنية على تحليل المحتوى الفراغي للألوان في الصور والتشكيلات الاتجاهية في الصور.

الدراسة الرابعة:

Chen-Hesin-Lliang, An Analysis of Image Qqueries and Iimage Retrieval Tasks in the Field of Art History, Advisor Rasmussen Edith.- University of Pittdsburgh, 1999. - 144p.- Ph.D.

تحقق هذه الدراسة في سمات أسئلة المستفيدين لطلبات الصور في مجال تاريخ الفن، من خلال مقارنة السمات المشتركة في الأسئلة بالاعتماد على دراسات سابقة متمثلة في دراسة "Enser and McGregor" عام ١٩٩٢ ودراسة Jorgensen عام ١٩٩٥ ودراسة Fidel عام ١٩٩٧. وتحاول اكتشاف أي سمات نوعية جديدة في مجال تاريخ الفن، وفحص العلاقات بين أغراض بحث المستفيدين ونوعيات طلبات الصور من خلال البحث في عناوين الصور والواصفات.

الدراسة الخامسة:

John Morris Zachary. An Information Theoretic Approach to Content Based Image Retrieval.-2000.- Ph.D.

تناقش هذه الدراسة ملامح اللون وما يخصه في استرجاع الصور والدور الذي يؤديه في الاسترجاع بالمحتوى. كما تناقش إمكانية التعبير عن اللون بالمصطلحات الكشفية، وأجرت الدراسة الاختبارات التجريبية على قاعدة بيانات تشمل ١٠ آلاف صورة تحاول من خلالها أن تناسب شكل الأسئلة المطروحة من جانب المستفيدين.

وكما يظهر من هذه الدراسات أنها انصبت على تحليل المكونات المادية للصور من حيث اللون والشكل والتركيب من خلال تقنيات معينة.. بينما اقتصرت دراسة العناصر الموضوعية للصورة على قياس أداء نظم الاسترجاع من خلال تحليل أسئلة عينة من المستفيدين في موضوعات الصور المتخصصة في تاريخ الفن، وأثر كل من مجالات علم النفس وعلم النفس المعرفي وعلم التكشيف وعلم الجمال في تكشيف قصص الخيال العلمي. ومن ثم لا يوجد دراسة تتعلق بتقييم أداء محركات البحث المتخصصة في الصور وإمكانية دعمها لخصائص اللغة العربية وهو موضوع هذه الدراسة.

أما عن الإنتاج الفكري العربي، فلم يتم تناول موضوع تكشيف وتحليل السمات الموضوعية للصور الرقمية من قبل، كما لم يسبق للرسائل الأكاديمية تناوله. ولا يوجد في هذا المجال سوى رسالة واحدة مسجلة هي:

- أحمد أحمد عبيد. نظم تكشيف الصور الثابتة في المؤسسات الصحفية المصرية: دراسة نظرية وميدانية. إشراف مصطفى حسام الدين. القاهرة: جامعة القاهرة، كلية الآداب، قسم المكتبات والوثائق والمعلومات، ٢٠٠٢م ماجستير.

وتتناول هذه الدراسة اتجاهات المؤسسات الصحفية في التعامل مع مجموعاتها من الصور الإعلامية التقليدية، إلى جانب أساليب التنظيم و المعالجة الفنية المختلفة التي تتبعها تلك المؤسسات مع مجموعاتها، وكيفية البحث والاسترجاع لتلك المجموعات، وأساليب الحفظ المادي للصور التقليدية.

إلى جانب دراستين عن محركات البحث هما :

١ - داليا نصار رياض. محركات البحث العربية على شبكة الإنترنت : دراسة تقييمية؛ إشراف محمود عفيفي، زين عبد الهادي. ـ جامعة حلوان : كلية الآداب، قسم المكتبات و المعلومات، ٢٠٠٤م - ماجستير.

تناولت هذه الدراسة تقييم محركات البحث العربية من خلال دراسة خصائصها البحثية وواقع استخدام اللغة العربية، والعلاقة بين محركات البحث ومناهج المكتبات،ودور محركات البحث في تقديم الخدمة المرجعية.

٢ - ضياء الدين عبد الواحد. واجهات الاستخدام لنظم استرجاع المعلومات المتاحة على شبكة الإنترنت : دراسة تقييمية؛ إشراف سهير أحمد محفوظ،زين عبد الهادي.-جامعة حلوان، قسم المكتبات والمعلومات، ٢٠٠٤م - ماجستير.

ب - قراءة في أدبيات الإنتاج الواردة في الدراسة:

إضافة إلى ما سبق، فإن هناك إنتاجًا فكريًا منشورًا صدر في بحوث ودراسات يتناول عددًا من الموضوعات ترتبط ارتباطًا كبيرًا بموضوع دراستنا هذه. وقد تباينت الموضوعات الفرعية في الحديث عن استرجاع المادة المصورة؛ إلا أن هذه العناصر المتباينة أخذت جميعا شكل التكامل للوصول إلى بناء نظم استرجاع معلومات للمادة المصورة الرقمية. حيث يأخذ الحديث عن تناول الإنتاج الفكري للصور الرقمية أكثر من شكل بحسب الجوانب التي تتناول الصور الرقمية. و يمكن ذكر بعض الجوانب مثل :-

- ✓ الصور الرقمية من منظور علم الحاسب في الاختزان و التكوين.
- ✓ الصور الرقمية في قياس آلات التصوير الرقمية و تقنية الكثافة.
- ✓ الصور الرقمية كنوع من أنواع التصميم الفني.
- ✓ الصور الرقمية التجارية " الدعاية و التسويق " .
- ✓ الصور الرقمية البحثية و التعليمية.
- ✓ الصور الرقمية في علم الاتصال.
- ✓ الصور الرقمية كأوعية للمعلومات المرئية " استرجاع المعلومات " .

ويهدف هذا الجانب التمهيدي من الدراسة إلى رصد الجوانب التي اهتم بها الإنتاج الفكري الخاص بالصور الرقمية كأوعية للمعلومات وعمليات التنظيم والبث للصورة الرقمية. وسوف يتم الاعتماد على قراءة تحليلية مبسطة للإنتاج الفكري الذي اعتمدت عليه الدراسة في البناء النظري لها.. إلى جانب التعرف إلى الاجتهادات العربية في هذا الموضوع وبيان مدى اهتمام الإنتاج الفكري العربي بتنظيم وبحث الصور الرقمية على الويب. وقد اعتمدت الدراسة على عينة قصدية من مصادر الإنتاج الفكري تتكون من مجموع ٧٥ مصدرًا حيث تتوافر بها البيانات البيليوجرافية كاملة، يمكن بالنظر إليها التعرف إلى الملامح الأساسية للإنتاج الفكري عن الصور الرقمية في بيئة الويب.

وتتبنى الدراسة في التحليل العناصر التالية :-

ب - ١ التوزيع الموضوعي للإنتاج الفكري للصور الرقمية.

ب - ٢ التوزيع النوعي للإنتاج الفكري للصور الرقمية.

ب - ٣ التوزيع الزمني للإنتاج الفكري للصور الرقمية.

ب - ٤ التوزيع الجغرافي للإنتاج الفكري للصور الرقمية.

ب - ٥ هيئات و مؤسسات الصور الرقمية.

ب - ٦ الإنتاج الفكري العربي عن الصور الرقمية.

ب - ١ التوزيع الموضوعي للإنتاج الفكري للصور الرقمية :-

تباينت الموضوعات الفرعية للحديث عن استرجاع الصور؛ إلا أن هذه العناصر المتباينة أخذت جميعا شكل التكامل للوصول إلى الصور النهائية لبناء نظم استرجاع معلومات الصور الرقمية. وقد جاءت مصادر الدراسة في ترتيب عددي يوافق ترتيب الموضوعات من الأهم إلى الأقل أهمية ؛ حيث يوضح الجدول التالي ترتيب الموضوعات بحسب إحصاء كل منها :-

فئات الموضوعات بحسب الترتيب العددي

الموضوع	عدد مصادر الموضوع
إستراتيجيات بحث الصور الرقمية	٢٥
تنظيم الصور الرقمية	٢٠
الوصف الفني لملفات الصور	١٥
تقنيات الصور الرقمية	٥
تطبيقات الصور الرقمية	٥
المعلومات المصورة	٥

☒ **الفئة الأولى :-**

اهتمت أكثر المصادر بعدد "٢٥" مصدرًا بموضوع إستراتيجيات بحث واسترجاع الصور من بالتعرض لدراسة كيفية عمل محركات بحث الصور وآليات البحث بها ؛ كدراسة Berinstein, paula. Turning visual : image search engines on the web, on line inc., 1998 التي تتحدث عن الخطوط العريضة للتعامل مع الصور الرقمية، وطبيعة أساليب البحث المتبعة مع الصور كوسيط لحمل المعلومات على الويب. أيضا دراسة أخرى هي Thomason, Larisa. Promotion Tip: Search Engines Get The Picture, 2002. و تهدف نماذج هذه الدراسة إلى عرض جوانب مخصصة لبحث و استرجاع الصور الرقمية دون باقي أنواع المعلومات المصورة، هذا فضلا عن الدراسات التي تناولت الآليات العامة لبحث مصادر الويب التي يمكن استخدامها مع تنظيم و بحث الصور الرقمية. وعملت دراسات إستراتيجيات البحث على تناول تعامل محركات بحث الويب مع ملفات الصور الرقمية و عمليات البحث في النص المصاحب لملفات الصور، غير

أن الباحث يرى أن هذه الدراسات قد تعاملت بعمومية محركات البحث بالتركيز على آليات بحث النصوص دون التأكيد على اختلاف ملفات الصور الرقمية. ثم دراسة Frequently Asked Questions - Picture Search. Fast Search & Transfer ASA, 2002 التي عملت على تبني مسح يشمل مجموعة من مستخدمي الصور ثم طرح أهم التساؤلات التي تتعلق ببحث واسترجاع الصور الرقمية من محركات بحث الصور على الويب. أيضا دراسة Notess, Greg. Searching beyond text : issues with multimedia searching, information today, 2000. وتمثل الدراسة أفضل نماذج تحليل استرجاع و بحث الصور الرقمية على الويب، حيث تعرض لنماذج استرجاع الصور بالكلمات المفتاحية والتصفح الموضوعي للمواد المصورة التي أضحت غاية مستخدمي الويب أكثر من بحث النصوص.[١] وهناك أيضا دراسة Larisa Thomason بعنوان Promotion Tip: Search Engines Get the Picture. وهي تهدف إلى عرض نماذج جوانب مخصصة لبحث واسترجاع الصور الرقمية دون باقي أنواع المعلومات المصورة. هذا فضلا عن تناولها الآليات العامة لبحث مصادر الويب، التي يمكن استخدامها مع تنظيم وبحث الصور الرقمية.[٢]

☒ **الفئة الثانية :-**

يختلف تنظيم الصور الرقمية بين مفهوم علم الحاسب الآلي وعلم المعلومات في أن الأخير يعتمد على تحويل وسيط الصورة إلى وسيط النص للتعبير عن المحتوى الموضوعي، ويتمثل ذلك في عملية التحليل الموضوعي للصورة بجانب الوصف المادي في تسجيلة واحدة.وهذا هو المعني به هنا في تنظيم الصور الرقمية.وقد أخذ تنظيم الصور الرقمية المكانة الثانية في اهتمام الإنتاج الفكري عن صور الويب بعدد " ٢٠ " مصدرًا. ويختلف الحديث بين إستراتيجيات البحث و تنظيم الصور في أن الأخير إنما يعنى به الباحث الحديث عن أساليب المعالجة الفنية الموضوعية دون الشكلية المادية، وتضمن الحديث عن تنظيم ملفات الصور عنصرين هما : استخدام قوائم رؤوس الموضوعات " اللغة المضبوطة" ، استخدام الكلمات المفتاحية الحرة " اللغة الطبيعية " . ويعكس اختيار أحد العنصرين السابقين الأسلوب المتبع في بحث و استدعاء الصور ؛ حيث يبنى على قوائم

---

(١) Notess, Greg. Searching beyond text: issues with multimedia searching, information today, 2000

(٢) Thomason, Larisa. Promotion Tip: Search Engines Get The Picture, 2002

المصطلحات استخدام المكانز أو القوائم العامة، كما يبنى على أسلوب الكلمات الحرة الويب بناء آليات بحث مثل استخدام البحث الحر. wild card ويعكس ذلك دراستان من الدراسات هما؛ الأولى دراسة Jake Cristine بعنوان: state of the arts : current applications for indexing images[1]، ودراسة Olivia Frost و Anna Noakes بعنوان Browsing Images Using Broad Classification Categories[2]. وقد تبنت الدراسة الأولى أسلوب التنظيم "التكشيف الحر" باقتباس الكلمات ذات الدلالة الموضوعية من النص المصاحب لملف الصورة أو من خلال الاعتماد على حقول HTML داخل بنية صفحات الويب، ومدى إمكانية استخدام ذلك في محركات البحث. أما الدراسة الثانية، فقد تبنت أسلوب الاعتماد على قوائم المصطلحات المرتبة منطقيًّا بحسب أنماط في خطط تصنيف المعرفة من العام إلى الخاص. واعتمدت بعض الدراسات التي سارت على هذا النحو بعض أشكال التصنيف العالمية مثل تصنيف مكتبة الكونجرس LCC أو تصنيف ديوي العشري DCC.

☒ الفئة الثالثة :-

جاء الحديث عن الوصف الفني لملفات الصور بعدد "١٥" مصدرًا كشق مكمل لموضوع تنظيم ملفات الصور، وذلك لإكمال الحديث عن المعالجة الفنية لأوعية الصور الرقمية. وتتركز دراسات الوصف الفني على معايير و أساليب الوصف الشكلي لعناصر ملفات الصور دون المحتوى الموضوعي منها مثل ؛ المؤلف و تاريخ التقاط الصورة و الحجم KBs و مقاس الصورة ... وغيرها. وعملت أيضا على محاولة الوصول إلى تسجيلات الصور الرقمية المعيارية بتكوين حقول الوصف المادي، فضلا عن مناقشة طبيعة الويب و طبيعة بناء التسجيلات الببليوجرافية داخل مختلف أنواع محركات بحث الصور المحددة " المخصصة" ومحركات بحث الويب. وتمثل دراسة Rusch-Feja Diann بعنوان: Metadata: Standards for Retrieving WWW Documents (and Other Digitized and Non-Digitized Resources أفضل ما قدم في هذا المجال وقد تناولت هذه الدراسة تحديدًا إظهار تحول المفهوم والمصطلح من

---

(١) Cristine, Jake. state of the arts: current applications for indexing images, 1999.

(٢) Frost, C. Olivia and Noakes, Anna. Browsing Images Using Broad Classification Categories, the University of Michigan, [2000?], cited 1/6/2002.

حقول الوصف الفني إلى حقول الميتاداتا التي استخدمت للدلالة على حقول الوصف الرقمية.[1] بينما ركزت دراسة DIG35: Metadata Standard for Digital Images على أحد معايير بناء الميتاداتا وهو معيار DIG35 لبناء التسجيلات الببليوجرافية للمادة المصورة الرقمية في بيئة الويب.[2] وعنيت بعض الدراسات بالتركيز على أحد معايير بناء الميتاداتا Metadata مثل دراسة DIG35: Metadata Standard for Digital Images. XMLCoverPages, 2002 في تبني معيار DIG35 لبناء التسجيلات الببليوجرافية للصور الرقمية في بيئة الويب. ويكتمل بذلك جانبا الحديث عن المعالجة الفنية للصور الرقمية.

☒ موضوعات أخرى:-

تعد الموضوعات الثلاثة السابقة أعمدة بناء هذه الدراسة؛ إلا أن الإنتاج الفكري للصور الرقمية لم يقتصر عليها فقط وإنما اشتمل على موضوعات أخرى مثل تقنيات الصور الرقمية على الويب وقد ذكرت الكثير من دراسات تنظيم ملفات الصور أهمية عرض الصورة و التحكم في المقاسات المختلفة أثناء عرض نتائج بحث محركات البحث. وركزت بعض الدراسات على تقنية درجة الكثافة للصور الرقمية Resolution، ومنها دراسة Reichmann, Michael H. Understanding resolution, luminous landscape, 2002 التي عنيت بأهمية هذه التقنية والتحكم بها في قواعد بيانات الصور. أيضا موضوعات مثل تطبيقات الصور في مختلف المجالات أو التركيز على تطبيق محدد مثل دراسة الخدمة الفنية الاستشارية للصور، وهي DIMP: The Use of Images to Support Instruction and Presentation. TASI Technical Advisory Service for Images, 2003 حيث تم التركيز في هذه الدراسة على استخدام الصور كنوع من عرض المعلومات المصورة و تقديم المعلومات. و قد تم التأكيد في الفصل الثاني على مثل هذه التطبيقات. يأتي في نهاية هذا العرض موضوع المعلومات المصورة حيث تناولت بعض الدراسات الصورة من المنظور العام كوعاء معلومات، ويمكن اعتباره الخلفية الأساسية لتاريخ المعلومات المصورة ثم دورها في الاتصال و نقل المعلومات. مثل ذلك دراسة Forger, Garry. Image, University of Arizona Library, 2002 ?, cited at 5/10/2003

---

(١) Rusch-Feja, Diann. Metadata: Standards for Retrieving WWW Documents (and Other Digitized and Non-Digitized Resources), Astronomical Society of the Pacific, 1998, cited at 2003

(٢) DIG35: Metadata Standard for Digital Images, XML CoverPages, 2002

التي يتضح من عنوانها الحديث عن الصورة الرقمية كوحدة جديدة من أشكال المعلومات.

ب - ٢ التوزيع النوعي للإنتاج الفكري للصور الرقمية :-

عملت الدراسة على مسح مصادر المعلومات التقليدية و الرقمية التي تتعلق بتنظيم وبحث الصور الرقمية في بيئة محركات البحث. وأثرت الخلفية التاريخية للإنتاج الفكري الذي بدأ فقط في النصف الثاني من عقد التسعينات على قلة توافر المصادر التقليدية في هذا الموضوع. وكان السواد الأعظم من مصادر المعلومات في الشكل الرقمي " مصادر الويب" . تمثلت مصادر الدراسة في التعامل مع مقالات الدوريات على الويب و الأوراق البحثية المقدمة في مؤتمرات مناقشة تنظيم مصادر الويب، أيضا المعلومات الإلكترونية في هيئات تنظيم الويب والصور الرقمية، والدراسات الأكاديمية النادرة التي تتعلق بتنظيم معلومات الصور.

واستعانت الدراسة ببعض قواعد البيانات المتخصصة في ذلك مثل: Dissertation Abstracts International UMI؛ إلا أن عمليات بحث هذه القواعد قد أشارت إلى ندرة الاهتمام بتنظيم الصورة كوعاء للمعلومات، في حين اهتم معظمها بما يتعلق بالصور الرقمية و التخزين على الحاسبات الآلية و ما يتعلق بها من تقنيات مثل: Content Based Image Retrieval؛ حيث استخدام الألوان و الشكل shape لحفظ وتنظيم الصورة الرقمية بعيدًا عن مفاهيم التكشيف أو المعالجة الفنية للمعلومات. ومن أمثلة الدراسات الأكاديمية دراسة Davis, Eric T.. A PROTOTYPE ITEM-LEVEL INDEX TO THE CIVIL WAR PHOTOGRAPH COLLECTION OF THE OHIO HISTORICAL SOCIETY, Advisor Zeng, Marcia Lei, Kent State University School of Library and Information Science,Master-thesis,1997. تعد الدراسة السابقة أقرب المصادر الأكاديمية لتنظيم الصور، إلا أنها قد اهتمت بالتنظيم التقليدي لأوعية معلومات الصور التقليدية في غير بيئة الويب.

ب - ٣ التوزيع الزمني للإنتاج الفكري للصور الرقمية :-

بدأت الصور الرقمية في الظهور في تاريخ يسبق تاريخ ميلاد الويب ؛ حيث تم إدخال الصور إلى الحاسب الآلي عام ١٩٦٥م.. بينما خرجت الويب إلى الضوء مع خروج متصفح موزايك NCSA Mosaic عام ١٩٩٣م. فضلا عن أن الدراسة لا تتناول

المواقع الأولى للويب وإنما تتناول مواقع محركات بحث الويب و هي فترة زمنية أخرى لاحقة لظهور الويب. إن هذا يعكس أن الإنتاج الفكري عن الصور الرقمية في بيئة الويب قد بدأ في النصف الثاني من عقد التسعينات. واعتمدت الدراسة في تحليل التوزيع الزمني لمصادر الويب على الجدول التالي :-

الترتيب الزمني لمصادر الدراسة

سنوات النشر	عدد مصادر الدراسة
١٩٩٦م	١
١٩٩٧م	٣
١٩٩٨م	٤
١٩٩٩م	٨
٢٠٠٠م	٧
٢٠٠١م	١١
٢٠٠٢م	١٧
٢٠٠٣م	١٦
٢٠٠٤م	٨

يشير الجدول السابق إلى الفترة الزمنية بين خروج الويب و بداية ظهور مشكلات تنظيم الصور الرقمية عليها التي بلغت ثلاثة أو أربعة أعوام. وتعكس أعداد مصادر الدراسة الواقع المنطقي للإنتاج الفكري عن الصور الرقمية وتضاعف أعداد ملفات الصور الرقمية على الويب ثم ظهور هذه المشكلة في مواقع بحث الويب ؛ حيث كان عام ١٩٩٦ أقل سنوات النشر في المصادر ثم أصبحت العلاقة طردية بين عدد المصادر و سنوات النشر كما يوضح الجدول السابق. يخالف ذلك عام ٢٠٠٤ حيث بدأت الدراسة في استعراض الإنتاج الفكري فقط دون التعرض للاقتباس من أي منه.

ب - ٤ التوزيع الجغرافي للإنتاج الفكري للصور الرقمية :-

شكلت المصادر الإلكترونية النسبة العظمى من مصادر الإنتاج الفكري لبناء هذه الدراسة. وعلى ذلك فإن الدراسة لم تعتمد على بيئة النشر التقليدية ذات أماكن النشر المحددة، وإنما اعتمدت على البيئة الموحدة الإلكترونية " الويب " . غير أن ذلك التوزيع الجغرافي مع التوزيع اللغوي القادم إنما يعكسان تفوق الإنتاج الفكري الأجنبي على نظيره العربي. وهذا إنما يأتي من أن ميلاد مصادر المعلومات الرقمية جاء على يد العالم الغربي أولا، يؤكد ذلك النشأة الأصلية للويب في وزارة الدفاع الأمريكية عام ١٩٦٩ ARPANET كمشروع للأمن القومي ثم تحول إلى الشبكة العالمية ثم ظهور الويب. و لم يعبر عن بيئة النشر العربية سوى القليل من المقالات مثل هانئ عطية في الحديث عن تكشيف الصور، ثم زين عبد الهادي في الحديث عن محركات بحث الإنترنت.

ب - ٥ هيئات ومؤسسات الصور الرقمية :-

اعتمدت الدراسة على الإنتاج الفكري التابع لبعض الهيئات الداعمة لتنظيم ملفات الصور الرقمية على الويب. وقد تعرضت الدراسة لهذه الهيئات في عنصر المعالجة الفنية للصور الرقمية حيث تختص بعض الهيئات بإنتاج المعايير الملائمة لأوعية معلومات الصور الرقمية، فيما يختص البعض الآخر بوضع معايير وصف مصادر الويب و من بينها المصادر المرئية. وأوضحت مصادر الدراسة أن أهم هيئات الصور الرقمية على الويب هي " TASI " Technical Advisory Service for Images، حيث تقف هذه الهيئة على أهم دراسات ومصادر الإنتاج الفكري حول تنظيم ومعالجة الصور الرقمية في بيئة الويب، خلاف الهيئات الأخرى التي تهتم فقط بتبعة معايير وصف و تحليل الصورة، هذا فضلا عن الاهتمام بتقنيات و آلات الصور الرقمية و عمليات تحويل الصور للشكل الرقمي، إضافة إلى تقديم أنواع التدريب المختلفة بحسب مرحلة التعامل مع الصور الرقمية. وتقترب الهيئات الأخرى في درجة الاهتمام بالصور الرقمية على الويب ويوضح الجدول التالي بعض هيئات الصور الرقمية :-

هيئات الصور الرقمية على الويب

الاسم الاستهلالي	اسم هيئات الصور
DC	Dublin Core
I3A	The International Imaging Industry Association
ELISE	Electronic Library Image Service for Europe
NISO	National Information Standards Organization
VADS	Visual Arts Data Service
VRA	Visual Resources Association
TASI	Technical Advisory Service for Images

يوضح الجدول السابق أن معايير وصف و معالجة الصور الرقمية عملت كنقطة تقابل بين هيئات تنظيم الويب عامة و المتخصصة في الصور الرقمية أو المواد المرئية. غير أنها تساوت جميعا في الحد الأدنى من الاهتمام بحصر وتنظيم الإنتاج الفكري عن الصور الرقمية. على الجانب الآخر، يوجد معهد أبحاث معلومات الصور الرقمية Institute for IMAGE DATA Research التابع لجامعة University of Northumbria at Newcastle؛ حيث يهتم بالأوراق البحثية حول معالجة و تنظيم الصور الرقمية.

ب - ٦ الإنتاج الفكري العربي عن الصور الرقمية :-

عكست مصادر الدراسة ضعف الإنتاج الفكري العربي في نشر مصادره عن البيئة الرقمية عامة وتنظيم الصور الرقمية خاصة. وقد اعتمدت الدراسة كلية على جمع الإنتاج الفكري الأجنبي باستخدام بيئة الويب وقد اقتصرت المصادر العربية النادرة على المقالات العامة عن محركات البحث وآليات البحث المستخدمة بها. عكست ذلك الدراسة التجريبية لزين عبد الهادي بعنوان " محركات البحث على شبكة الإنترنت " واهتمت هذه الدراسة فقط بالآليات العامة لمحركات بحث الويب دون التطرق من قريب أو بعيد إلى تنظيم و بحث الصور الرقمية.(١) وسارت الدراسات الأخرى على الدرب نفسه؛ كدراسة خالد محمد رياض عن المقارنة

(١) زين عبد الهادي. محركات البحث على شبكة الإنترنت: دراسة تجريبية مقارنة، **مجلة المكتبات والمعلومات العربية**، ع، أبريل ٢٠٠٢.ص ١٠.

العامة بين أدلة و محركات البحث.[١] إلى جانب المقالة الرقمية عن البحث في محركات البحث لمساعد بن صالح الطيار.[٢] إلا أن دراسات أخرى قد اهتمت بتنظيم ومعالجة مصادر الويب مثل دراسة هانئ محيي الدين عطية بعنوان جهود تصنيف المعرفة في البيئة الإلكترونية، واهتمت الدراسة بتتبع أنواع تنظيم وتصنيف المعرفة على الويب لوضع شكل يلائم مصادر الويب.[٣] يضاف إلى ذلك دراسة جوانب تكشيف الصور لهانئ محيي الدين عطية بعنوان: تكشيف الصور بين العلم والتعلم التي عملت على إبراز أهمية تدريب الطلاب على التعامل مع محتوى الصور، ومحاولة التعرف إلى المعاني المتباينة داخل الصورة الواحدة. وكان ذلك في عرضه لمحتوى عام يشمل القضية الفلسطينية.[٤] وقد تناولت الدراسة العوامل التي تقف وراء تباين اختلاف تكشيف الصور من بيئة إلى أخرى.

عاشرًا محتويات الدراسة :-

تتكون الدراسة من خمسة فصول بجانب عرض المقدمة وملحق الرسالة، وكان تتابع الفصول كالتالي:-

الفصل الأول تعرض لأنواع الأدوات الأساسية لبحث الويب من محركات وأدلة بحث بجانب قواعد بيانات الويب، إلى جانب تقسيم مقترح لأنواع محركات البحث على الويب وصولا إلى محركات بحث المادة المصورة والحديث بتركيز أكثر على نظم استرجاع الصور الرقمية.

الفصل الثاني تناول الحديث عن أوعية المعلومات المرئية وتبنى هذا الفصل شكلًا محددًا لتقسيم أوعية المعلومات المرئية على الويب والحديث عن هذه الأنواع انتهاء بالصور الرقمية الثابتة بموضوعاتها العلمية والإعلامية والعامة.

---

(١) خالد محمد رياض. **أدلة ومحركات بحث شبكة الإنترنت: دراسة مقارنة** -. الاستراتيجية العربية الموحدة للمعلومات في عصر الإنترنت ودراسات أخرى، الاتحاد العربي للمكتبات والمعلومات، تونس، (أكتوبر١٩٩٩). ١٣٩ ص.

(٢) مساعد بن صالح الطيار. **محركات البحث العربية وطرائق البحث فيها**. - الجزيرة، ٢٠٠١، تاريخ الاستشهاد ١٠/١٠/٢٠٠٢، مسار http://www.al-jazirah.com.sa/evillage/22112001/wr1.htm

(٣) هانئ محيي الدين عطية. جهود تصنيف المعرفة في البيئة الإلكترونية: دراسة استكشافية -.**الأكاديمية للمكتبات والوثائق والنظم المعلوماتية**، ع١ (٢٠٠٤)-. ص ص٥-٧٧.

(٤) هانئ محيي الدين عطية. **تكشيف الصور بين العلم والتعلم: ورقة في ملف القضية الفلسطينية** .- أسيوط: المؤتمر القومي الخامس لاختصاصيي المكتبات والمعلومات والأرشيف في مصر، ٢٠٠١.

الفصل الثالث تناول معالجة الصور الرقمية وذلك في قسمين تناول في القسم الأول منه المعالجة الفنية للصور الثابتة، ومحاولة التطرق إلى بعض المعايير التي تتبنى تحليل ووصف الصور الرقمية مع التركيز على معيار دبلن كور بعناصره المختلفة، وتناول القسم الثاني الحديث عن المعالجة الآلية لخصائص اللغة العربية وأثرها على تشكيل آليات الاستدعاء والبحث على الويب.

الفصل الرابع قام بدراسة تجريبية وذلك بالتطبيق على عينة غرضية من محركات بحث الصور الثابتة، واختبار مجموعة من خصائص هذا النوع التي تدور في مجملها حول تنظيم واستدعاء الصور الثابتة، بالإضافة إلى اختبار بعض آليات بحث اللغة العربية التي اقتصرت في اختبارها على محرك بحث Google.

الفصل الخامس عمد إلى دراسة طبيعة ترشيح المعلومات المصورة على الويب/ ودراسة الخصائص التي يتحلى بها مرشح الصور الرقمية داخل محركات البحث العربية، فضلا عن صياغة المراحل المقترحة لعمل مرشح للصور بما يوافق البيئة العربية والإسلامية.

الفصل السادس اقتصر على عرض المواصفات المقترحة لبناء محرك بحث صور ثابتة يدعم خصائص اللغة العربية، وذلك ابتداءً من عمليات اختيار ملفات الصور الرقمية مرورًا بتنظيمها ثم آليات البحث والاستدعاء للصور باللغة العربية يمكن من خلالها تحقيق أفضل درجة تحقيق لنتائج البحث أمام المستفيدين.



الفصل الأول

البحث على الويب
مفاهيمه وأساليبه وأنواعه



البحث على الويب
مفاهيمه وأساليبه وأنواعه

٠/١ التمهيد

إن الإنترنت بوصفها مستودع المعلومات الإلكترونية للعالم قد دخلت في تطبيقات عدة، واتسع مجالها حتى وسع كل موضوع من موضوعات المعرفة، مما جعل من عملية تنظيمها أمرًا شاقًا على كثير من المتخصصين. فلقد ضمت الإنترنت معلومات في المجالات كافة وبكل أشكال المواد المسموعة والمقروءة والمصورة، وهذا الخضم الهائل متعدد الموضوعات مختلف الأشكال يحتاج إلى طرق بحث وأساليب مختلفة تبدأ بأبسط الطرق وهي التصفح وتنتهي بآليات استرجاع متطورة. ومع تزايد مواقع الإنترنت والتي يقدر عددها بـملايين المواقع، ومئات أخرى تضاف كل يوم، ومع إضافة أنماط جديدة من التقنيات تصبح عمليات البحث والاسترجاع التي تعتمد في محورها على استخدام محركات البحث أو أدوات البحث في غاية الأهمية.

إن محور هذا الفصل هو الحديث عن أدوات البحث على الإنترنت وأنواعها المختلفة، ثم الحديث عن المادة المصورة كوسيط معلومات تضمه الإنترنت وهو محور هذه الدراسة ككل، ثم انتهاء بالفارق بين استرجاع المادة المصورة واسترجاع النص.

تعريف الإنترنت

بعيدًا عن الخوض في نشأة الإنترنت وتطورها، فإن الحديث سوف يكون منصبًا على ماهية الإنترنت من المنظور المعلوماتي، ولطالما تعددت تصنيفات الإنترنت من هذا المنظور بين مرجع، وقاعدة بيانات، وشبكة معلومات. ويقترح

الباحث تعريفًا إجرائيًا بأن تصنف الإنترنت على أنها مصدر معلومات يتم اقتناؤه في المكتبات تحت مسمى خدمة الإنترنت وإتاحة هذا المصدر كما تتطلب طبيعة الخدمة في قاعات مخصصة لذلك وتقديم طرائق مختلفة في التعامل مع المستفيدين."

١/١ الويب

إن هوية الإنترنت الحقيقية التي توارت خلف بريقها ولمعان عرض المعلومات المبهر في وسائط سمعية وبصرية ونصية، ليست هي الويب، فالأخيرة ليست سوى جزء من الإنترنت يتكامل مع غيره من الأجزاء (جوفر، آركي، وايس، بروتوكول نقل الملفات، وتيلنت) التي تكون في مجموعها الإنترنت أو الشبكة العالمية. والويب هي الجزء المعني بالوسائط المتعددة والروابط الفائقة القائمة على استخدام لغة الترميز المعيارية، وهي ما يعنينا في هذه الدراسة.

وفيما يلي تعريف بمصطلحات الويب كما سيتم التعامل معها في هذه الدراسة :-

١/١/١ بنية الويب ومكوناته

إن الويب كما سلف الذكر هي الجزء الأكثر شهرة واستخدامًا بين مستخدمي الإنترنت. وهذا الجزء له بنية خاصة وتكوين محدد يتم استخدامه في عرض المعلومات واسترجاعها من مستودعاتها. وتعتمد الويب على مكونات أساسية في تكوين هذه البنية ذكرتها Cheryl Gould كما يلي[١]:

- **الرابط** : هو أداة للوصل بين صفحتين من صفحات الويب أو وسيط من وسائط المعلومات أو وسيط آخر داخل مواقع الويب مثل ربط الكلمات المفتاحية بالمادة المصورة المعبرة عنها.
- **الصفحة الدليلية**: وهي الصفحة أو الموقع الذي يصل إليه متصفح الإنترنت حالة الدخول إلى الويب وفتح برنامج تصفح الإنترنت، وتكون هذه الصفحة مسجلة من قبل بواسطة المستخدم.

---

(١) غولد، تشيرل. البحث الذكي في الإنترنت: أدوات وتقنيات للحصول على أفضل النتائج. ـ الرياض: مكتبة الملك فهد الوطنية، ٢٠٠١م.١٤٨ ص.

صفحات الويب : هي الوحدات التي تتكون منها المواقع، وهي تعد الوحدة الصغرى من الوحدات المكونة للويب، وهي تحتوي على معلومات متعددة الأشكال قد تكون مرئية أو مسموعة أو مجرد نصوص وفقرات.

لغة الترميز المعيارية للنص المترابط : وهي اختصار للمصطلح Hypertext Markup Language، وهي إحدى لغات البرمجة التي تستخدم في تصميم صفحات ومواقع الويب. وما يمكن مشاهدته للدلالة على ذلك هو وجود مصطلح html أو htm في نهاية امتداد كل عنوان من عناوين الويب .URL

متصفح الإنترنت: وهو البرنامج المستخدم مع نظم تشغيل النوافذ windows على مختلف إصداراتها، ومن خلاله يتم قراءة ملفات الإنترنت وزيارة كل مواقعها. ومن أهم البرامج المستخدمة في ذلك هي internet explorer, Netscape .navigator

موقع الويب : اسم لمجموعة من الصفحات مكتوبة بلغة html ومعتمدة على الروابط الفائقة التي لا تكون مرتبطة فقط في التصميم وإنما أيضا في وحدة الموضوع.

وتعتبر المكونات السابقة هي التطبيقات أو الأدوات التي تعمل على الويب، حيث تعتمد الويب إضافة إلى ذلك على المكونات المادية مثل الخادمات ووسائل الاتصال التقليدية والمتقدمة.

ويضيف الباحث لما سبق ما يلي:

مسار الموقع URL: ويطلق على مسار مواقع الويب اختصار URL الدال على مصطلح محدد المصادر الموحد. ويتألف مسار مواقع وصفحات الويب من مجموعة من الأجزاء تتكامل في وحدة واحدة، ولكل جزء منها مفهوم محدد يساعد في تخصيص المعلومات وفق ما تكون عليه في الموقع. ويعد مكان المسار رابطًا للخادم الذي يحوي صفحة الويب المطلوبة، ويتجلى في مسار مواقع الويب الدليل على صفحة الويب كجزء من الإنترنت؛ حيث تأتي حروف www المعبرة عن الويب مصطلح world wide web، بعد حروف http المعبرة عن بروتوكول تعامل الإنترنت Hyper Text Transfer Protocol.

ويتكون محدد الويب في أدنى حالاته من قسمين فقط مثل http:// www.google.com الذي يتكون من قسمين أو لهما هو بروتوكول التعامل مع الويب http، أما الثاني فهو مجال "نطاق" الموقع Domain name مثل www.google.com. ومن الممكن أن يشمل المسار أجزاء أخرى مثل مسار /http://www.usps.gov/consumer pats.htm، فإن الجزء الثالث بعد البروتوكول والمجال "النطاق" يسمى directory وهو في العنوان السابق consumer. أما الجزء الأخير من المسار يسمى الموضوع المخصص specific file وهو في المثال السابق pats. وعلى هذا فإن محدد المصادر يمكن أن يشتمل على أجزاء عدة بحسب تخصص الصفحة التي يريد المستخدم الوصول إليها.

وتختلف البروتوكولات في مسارات الإنترنت حسب المعلومات التي تقدمها هذه المواقع، وهي كما يلي:-

الجدول رقم (١-١) يوضح أنواع بروتوكولات الإنترنت

الخدمة	البروتوكول
بروتوكول نقل النص الفائق	http:// Hyper Text Transfer Protocol
بروتوكول نقل الملفات	ftp:// File Transfer Protocol
بروتوكول البريد الإلكتروني	mailto: Mail Protocol
بروتوكول مجموعات الأخبار	News: News Protocol
بروتوكول بحث الملفات	Gopher: gopher protocol
بروتوكول الربط عن بعد	Telnet: telnet protocol

واستكمالا لتحليل مسار مواقع الويب، فإن الشق الثاني وهو المجال "النطاق" Domain يتكون من عنصرين أو لهما: التعريف بالويب www، وثانيهما اسم الموقع مثل Google في www.google.com، ثم تليه الحروف الدالة على نوع الموقع مثل com في المثال السابق. وتختلف الحروف باختلاف أنواع مواقع الويب كما يلي :-

الجدول رقم (١-٢) يوضح أنواع صفحات الويب

النوع	الاختصار	الاسم
مواقع تجارية	com	Commercial
مواقع تعليمية أكاديمية	edu	Educational
مواقع حكومية	gov	Governmental
مواقع عسكرية	mil	Military
مواقع للشبكات	net	Networks
الهيئات غير الهادفة للربح	org	For non profit

- **الميتاداتا :** ارتبط مفهوم الميتاداتا بالاستخدام مع البيئة الرقمية للتعامل مع وصف مصادر الويب. وهناك تعريفات عدة للميتاداتا نسرد منها تعريف Steven J. Miller الذي تناول الميتاداتا من حيث المعالجة الفنية والتنظيم لمصادر الويب بأنه "الميتاداتا = البيانات عن البيانات : أي أنها حقول بيانات رقمية يتم تنظيمها لتكشيف وتجميع ووصف وتنظيم المعلومات على الخط المباشر ومصادر المعرفة"[١]. وتعريف Chris Taylor حيث عرف الميتاداتا على أنها " بيانات معيارية تم بناؤها لوصف خصائص المعلومات الرقمية، وهي تتشابه في خصائصها مع الفهرسة داخل المكتبات "[٢]. ويعرف الباحث الميتاداتا بأنها " مجموعة من حقول البيانات الرقمية الهادفة لوصفة وتنظيم المصادر الرقمية".

١/١/٢ البحث داخل الويب

إن الحصول على المعلومات الصحيحة هو غاية عملية البحث، ولكن هذه العملية قد تختلف في درجاتها من حيث الكيفية (طريقة البحث) والأداء (نتيجة الاسترجاع). فمسألة البحث على الويب لها طرق عدة، ويعتمد الباحثون في اختيار

---

(١) Miller, Steven J. Metadata and Cataloging Online Resources, University of Wisconsin-Milwaukee Libraries, 2003, Visited at 10/5/2004, http://www.uwm.edu/~mll/resource.html

(٢) Taylor, Chris. metadata, University of Queensland Library, 2003, Visited at 10/5/2004, http://www.cybrary.uq.edu.au/iad/mainmenu.html

وسيلة البحث لاسترجاع معلومات الويب على مستوى مهارة المستفيد في البحث واستخدامه للويب. وتحتاج كل وسيلة لبحث الويب إلى أدوات مخصصة مثل الاعتماد في المراحل المتقدمة على استخدام أدلة ومحركات البحث.

وقد أوضحت الباحثة Laura Cohen ست وسائل أساسية لإيجاد المعلومات داخل الويب وذلك في مقالها الذي جاء بعنوان how to find information on the internet وهو الذي يعتبر في رأي الباحث عنوانا أعم من المحتوى المقصود؛ لأن محتوى المقال كان قد تركز على البحث داخل الويب وليس الإنترنت كما أشار العنوان. وتضمنت هذه الوسائل طرائق الوصول السابقة مع أخرى جاءت كما يلي:-

١. الاتجاه مباشرة إلى موقع المعلومات عن طريق استخدام مسار محدد للمعلومات "URL".

٢. التصفح باستخدام برنامج تصفح الإنترنت.

٣. الانصال بإحدى مجموعات الأخبار أو مجموعات المناقشة على الويب.

٤. استكشاف المعلومات غير المرئية داخل الويب من خلال بحث قواعد البيانات.

٥. تصفح أحد أدلة البحث على الويب.

٦. الاتجاه إلى أحد محركات البحث.

غير أن وسائل البحث السابقة تعتمد في الوصول إليها على خبرة المستفيدين في البحث، وإذا ما أخذنا في الذهن صفة غير المتمرسين في بحثهم الويب، فإننا نتجه لوصف الجزء الأكبر من مستخدمي الويب، مما يجعل طرق البحث باستخدام أدوات البحث أو عنوان الموقع هي الطرق الأكثر استخداما في الوصول إلى المعلومات. ويمكن بيان كيفية استخدام وإجراءات كل طريقة من طرائق البحث السابقة كما يلي :-

١/٢/١/١ استخدام مسار أو عنوان موقع المعلومات (URL)

إن استخدام هذا الأسلوب يتأتى من معرفة المستخدم لعنوان أو مسار موقع المعلومات المطلوبة على الويب، وذلك بكتابة عنوان الموقع المطلوب في الفراغ المخصص لكتابة اسم الموقع تماما كما هو؛ إلا أن من أهم خصائص هذا الأسلوب هو أن يكون الباحث ذا معرفة بالعنوان أو المسار الصحيح للموقع، مع

مراعاة كل العلامات والفواصل وتغير أشكال الحروف small or capital letters، إذ أن خطأ بسيطا في أي حرف من حروف المسار قد تؤدي به إلى موقع آخر أو إلى صفحة النتائج غير الصحيحة التي تشير إلى عدم وجود هذا المسار. فعلى المستفيد حفظ مسارات مواقع الإنترنت بشكل صحيح وكتابتها بدقة.

١/١/٢/٢ استخدام برامج تصفح الإنترنت Web Browsers

تحتاج صفحات ومواقع الويب بطبيعتها إلى بيئات خاصة تتيح لها أن تعمل وتسمى هذه البرامج بمتصفحات الإنترنت. ويقصد بها "البرامج الخاصة لعرض صفحات ووثائق الإنترنت، أو تلك البرامج التي تعمل في بيئة الإنترانيت Intranet وهي تقنية الربط باستخدام شبكة الإنترنت. ويوجد الكثير من برامج تصفح الويب المستخدمة حاليا، منها المستكشف Internet Explorer، ونيتسكيب Netscape Navigator، وأو برا Opera، وموزايك NCSA Mosaic.

ويعد البرنامجان الأول والثاني الأكثر استخدامًا بين المستفيدين في المجموعة السابقة. أما الأول فأنتجته شركة مايكروسوفت Microsoft، وقد خرج لهذا البرنامج إصدارات متعددة بدأت مع إنتاج نظام التشغيل windows95، وكان آخرها Internet Explorer ver.6. أما البرنامج الثاني وهو Netscape Navigator فأنتجته شركة Netscape وكانت آخر إصداراته هي Netscape 8.1. وتستعد الشركات المنتجة لهذين البرنامجين لإخراج الإصدارات القابلة للاستخدام بالصوت speech-enabled browsers.

وتعتمد برامج التصفح على جانبين فقط في الاستخدام؛ الأول في عرض صفحات ومواقع الويب، وهو الجانب الأكثر استخدامًا تبعًا لطبيعة تلك البرامج؛ حيث لا تخصص قدرات لبحث صفحات الويب سوى استخدام فراغ العنوان وهو الجانب الثاني لاستخدام المتصفحات والذي يعتمد بصفة أساسية على معرفة عنوان صفحة الويب URL مسبقًا. وتتيح المتصفحات من خلال واجهاتها التنقل بين روابط الصفحات المختلفة بحيث يمكن التحرك من صفحة إلى أخرى؛ إلا أن هذا يعتمد بطبيعته على وجود رابط الصفحة التي سيتم الانتقال إليها، وهو ما يعد نقصًا كبيرًا في البحث باستخدام المتصفحات. وفي حالة استرجاع المادة المصورة تبعًا لهذا الأسلوب فإنه من اللازم توافر لقطات المادة المصورة أو وحدات أصغر للوصول إلى روابط وحدات المادة المصورة الرقمية في الإطار الأصلي لها.

١/١/٢/٣ استخدام مجموعات الأخبار على الويب News Groups

يعتمد استرجاع المعلومات هنا على استقبال رسائل البريد الإلكتروني والتي تقدم المعلومات الجارية لما يخص الموضوع المحدد؛ حيث تحيط هذه الرسائل الباحثين بالمقالات حديثة النشر أو ما تم نشرها بين أعضاء المجموعة. ويذكر T. Matthew Ciolek أن مجموعات الأخبار قد سارت على نسق خدمات البريد الإلكتروني ومجموعات الدردشة في نمو المجتمعات الإلكترونية، وتستقطب مجموعات الأخبار المستفيدين الذين يشتركون في السمات العلمية والبحثية. وتتعدد الأنشطة داخل هذه المجتمعات من تبادل الآراء والتراسل ونشر الأبحاث الحديثة في الموضوع[١].

وتحتوي مجموعات الأخبار بدورها على قواعد بيانات يمكن للباحثين إجراء البحث بها واسترجاع التقارير أو المقالات السابق تخزينها على قاعدة بيانات. ويتطابق البحث داخل قواعد البيانات مجموعات الأخبار مع قواعد البيانات الأخرى على الويب، ويخضع البحث بدوره إلى أساليب بحث محددة لقاعدة البيانات، غير أن محتوى تلك القواعد يتوقف على ما تضمه مجموعة الأخبار في هذا الموضوع.

١/١/٢/٤ بحث قواعد البيانات "المعلومات غير المرئية"

تعتبر قواعد البيانات على الويب الجزء غير المرئي أو الخفي من الويب، وتأخذ قواعد البيانات مصطلحات مختلفة مثل Invisible Web أو Hidden Web أو Deep web. وقد أخذت قواعد البيانات تلك المصطلحات لاختفاء مصادر معلوماتها عن التقنيات التي تستخدم في البحث على صفحات الويب المرئية، مما يعني عدم إمكانية الوصول إليها وتكشيفها. فبينما استرجاع المعلومات على الويب المرئية يتم من خلال محركات البحث التي يستخدم فيها برنامج يسمى الزاحف Crawl، يتم استرجاع المعلومات داخل قواعد البيانات "الويب غير المرئية" بالوصول المباشر إلى صفحة الموقع الذي يحوي قاعدة البيانات ثم البحث داخلها.

---

(١) Ciolek, T. Mat:thew. Internet Structure and Development: On Strat:egic Uses of the Archetypes of the Networked Mind, author, 1999, Visited at: 10/10/2003,: http://www.ciolek.com/PAPERS/pnc-taipei-99.htm

ولقد ذكر Paul Pedley أن مصطلح الويب غير المرئية يمكن أن ينطوي على أربعة مستويات من اختفاء مصادر قواعد البيانات، وقد حدد هذه المستويات اعتمادًا على درجة الوصول إلى قاعدة البيانات. وهي كما يلي[١]:-

١/١/٢/٤/١ الويب غير الشفافة

وهي تشمل قواعد البيانات كبيرة الحجم على الويب، التي تتكون بدورها من مجموعات كبيرة من مواد المعلومات، وعلى الرغم من ملاءمة هذه المواد لعمل برنامج الزاحف وإمكانية اقتنائها داخل قواعد محركات البحث؛ إلا أن برامج الزاحف لا تستطيع الوصول إليها ويرجع Paul Pedley هذا إلى عوامل مختلفة منها:-

عمق الزحف: تعمل محركات البحث على زيادة عمق التحليل والبحث لبرامج الزاحف لتحقيق أفضل أشكال الاقتناء، خاصة في الوقت الذي تتزايد فيه أعداد صفحات مواقع الويب بشكل مستمر. وتعمق برنامج الزاحف داخل موقع الويب وتحليل كامل صفحاته لأجل جمعها من شأنه رفع تكاليف العمل داخل محركات البحث. وتتجنب محركات البحث إظهار مدى العمق الذي تعمل به برامج الزحف داخل صفحات الويب.

تأخر التحديث الدوري: فإضافة صفحة جديدة إلى محرك البحث يحتاج إلى وقت قد يصل إلى شهر أو أكثر حتى يتم تمييزه ببرنامج الزاحف، وعلى الجانب الآخر فإن الرجوع إلى مواقع سابقة للبحث عن التحديثات الأخيرة بها من صفحات مضافة يحتاج إلى وقت أطول من الوقت الذي يستغرقه التحديث. وعلى ذلك فإن جزءًا من مواد المعلومات المحدثة داخل مواقع الويب التي تم إضافتها مسبقًا سيظل بعيدًا عن أيدي الباحثين فترة من الوقت، وهو ما يحدث في النوع الحالي من قواعد البيانات.

ازدياد النتائج المستدعاة Maximum Number of Viewable Results: تتسم محركات البحث بكبر حجم قواعد بياناتها، ويتضح ذلك في أعداد النتائج المسترجعة داخل قاعدة البيانات التي قد يصل عددها إلى ملايين الصفحات، مما يعني معه أن محركات البحث تعمل على الإهمال المتعمد لاستبعاد النتائج الأقل

---

(١) Pedley, Paul. The Invisible Web, The Library Associat:ion, 2000, Visited at: 20/10/2003,: http://www.la-hq.org.uk/directory/record/r200011/article2.html

علاقة بالنسبة لكلمات البحث ومثل هذه النتائج تظل دائما الجزء غير المرئي أمام الباحثين.

زيادة الروابط غير النشطة: وهذا بدوره لا يرجع إلى عمل محرك البحث وإنما خطأ في تفعيل روابط بعض الصفحات التي لا يستطيع برنامج الزاحف الوصول إليها سوى باستخدام مسار الصفحة URL أو الإضافة المادية عن طريق مصممي صفحات الويب.

٢/٤/٢/١/١ الويب الخاصة

وهي تتكون من مجموعة من صفحات الويب القابلة للتكشيف بواسطة برنامج الزاحف، غير أنها تستخدم حالة من حالات ثلاث تالية لمنع دخول برنامج الزاحف إليها. وهذه الحالات هي:-

استخدام كلمات المرور للتعرف إلى محتوى الصفحة.

تهيئة ملف robots.txt المحدد لعمل برنامج الزاحف لمنع التعامل مع صفحة الويب.

استخدام حقل noindex meta tag للتوقف عن إضافة هذه الصفحة.

٣/٤/٢/١/١ الويب ذات الملكية

ويقصد بهذا النوع مجموعة قواعد البيانات التي تتوقف على وجود حساب أو ربط Registration مع موقع القاعدة. وعلى الرغم من أن هذا النوع يقدم خدمات مجانية، فضلا عن سهولة إتمام عملية الربط بالقاعدة؛ إلا أن أفضل برامج الزاحف حتى الآن لا تستطيع التعامل مع إقامة ربط لقواعد البيانات. مما يجعل من الصعب عبور الزاحف إلى محتوى هذه القواعد.

٤/٤/٢/١/١ الويب غير المرئية

ويعد هذا الجزء من الويب هو الواقع الفعلي لمفهوم الويب غير المرئية. ويمكن لأي جزء من صفحات الويب أن ينجلي أمام الباحثين إذا ما امتدت إليه برامج محركات البحث لتكشيفه وإخضاعه للبحث أمام المستفيدين، ويختلف الجزء الحالي من الويب في البناء الفني عن الشكل المستخدم لدى برامج الزاحف فلا يستطيع برنامج الزاحف استخدام محتوى تلك الصفحات. ويشير Paul Pedley إلى أن

ذلك يتوقف على احتواء صفحات الويب على أنواع من الملفات لا تتعامل معها محركات البحث مثل الملفات المضغوطة وملفات pdf وملفات الفلاش Flash، ولا يكفي الجزء البسيط من حقول الميتاداتا المصاحب لهذه الملفات في إعانة برامج الزاحف على تحليلها وتكشيفها التي تحتاج إلى التعامل مع كامل المتن لصفحة الويب في تكشيفها وبحثها.

٥/٢/١/١ تصفح أحد أدلة البحث على الويب

يطلق على هذا النوع من أدوات البحث مسميات عدة مثل subject guides internet resource guides, internet directories, internet indices, internet catalogs وكلها تعطي معنى القوائم أو الموضوعات المقسمة. وتعتمد أدلة البحث على تقسيم المعرفة أو العلوم إلى أقسام عامة ثم إلى أقسام أكثر تخصصًا أو قوائم فرعية، ومن ثم يمكن الحصول على الموضوع ومن ثم المواقع التي ترتبط مباشرة بهذا الموضوع المخصص. وقد تعددت الأدلة على الويب ولكن يمكن تقسيمها إلى الفئات التالية:-

١. أدلة بحث موضوعية: وهي تلك التي تعتمد على التقسيمات العلمية للموضوعات وهي بدورها تنقسم إلى نوعين: أدلة البحث العامة، وأدلة البحث المتخصصة؛ والعامة هي التي تغطي مجالات المعرفة كافة، وهي قد لا تعتمد على تقسيم موضوعي محدد وإنما تعتمد على تقسيم عام لكل المعرفة وقد يتضمن هذا التقسيم موضوعات خاصة بالأسرة أو الرحلات أو الترفيه. أو قد تعتمد على خطة تصنيف عالمية، أو خطة تصنيف قومية، أو متخصصة. أما أدلة البحث الموضوعية المتخصصة فقد تغطي موضوعات متخصصة مثل مجال المكتبات، أو الطب مثلا.

٢. أدلة بحث نوعية: وهي تلك التي تقصد فئة معينة من المستفيدين مثل أدلة مواقع الأطفال أو المسنين، أو ذوي الاحتياجات الخاصة، أو المرأة أو غيرها وهي بدورها قد تعتمد على خطة تصنيف عالمية، أو متخصصة.

٣. أدلة بحث جغرافية: وهي تلك التي تعتمد على تغطية المواقع التي تنتمي إلى دولة معينة أو منطقة جغرافية تتحدث بلغة واحدة.

وتعد الهيئات التجارية هي أغلب الهيئات التي تعمل على وضع أدلة البحث بجميع أنواعها وذلك بهدف وضع الإعلانات، حيث إنها الأكثر زيارة واستخدامًا على الويب.

ولقد وضع Paul Pedley مجموعة من العناصر التي يتم عندها الاستعانة بأدلة البحث، إذا ما أراد الباحث الحصول على معلومات قيمة في مجاله وهي :-

١. عندما تكون هناك فكرة أو موضوع واضح في ذهن الباحث.
٢. عندما يريد الباحث حصرًا بالمواقع في موضوع واحد.
٣. عندما يريد الباحث الحصول على قائمة بالمواقع التي تخص موضوعًا مخصصًا لا يتخللها شوشرة من المواقع التي لا تتصل بالموضوع.
٤. عندما يريد الباحث الحصول على عناوين أو كلمات مفتاحية بغض النظر عن الحصول على مواقع أو مصادر تتصل بالموضوع.
٥. عندما يريد الباحث تجنب عرض تلك الصفحات التي تحتوي على محتوى قليل عن الموضوع، والتي غالبًا ما تكون ضمن نتائج محركات البحث.

٦/٢/١/١ الاتجاه إلى أحد محركات البحث

تعمل محركات البحث بالاعتماد على برنامج يطلق عليه الزاحف أو الديدان أو العنكبوت أو الإنسان الآلي،وهو البرنامج الذي يسبح في فضاء الويب ويتعرف إلى المواقع المختلفة لتحديد ماهية هذه المواقع ثم ضمها إلى قاعدة بيانات أداة البحث المنتمي لها البرنامج. ويأخذ الزاحف أو العنكبوت طريقه اعتمادًا على تزويده بآليات لاكتشاف صفحات الويب، متخذًا من بنية ملفات الويب المكتوبة بـ html أداة لتحليل الصفحات وتكوين الروابط الموصلة لها فيما بعد. ويبدأ برنامج الزاحف عمله في مسح فضاء صفحات الويب اعتمادًا على إستراتيجيات تحرك مختلفة؛ أولها يتمثل في الاعتماد على قائمة لمجموعات من المواقع مرتبة زمنيا داخل خادم أو أكثر من خادمات الويب وتعرف بقائمة " الحديث على الويب " ، أما الشكل الثاني فهو الاعتماد على قائمة موضوعة يدويًا من مصممي صفحات الويب بأهم المواقع التي يمكن أن ينطلق منها برنامج الزاحف، والشكل الثالث هو الاعتماد على قوائم مجموعات المناقشة أو الأخبار في المواقع التي تضمها لتزويد البرنامج بها.

ويرى Paul Pedley أنه على الرغم من وجود فارق كبير بين محركات البحث والأدلة البحثية في عناصر عدة مثل البحث الخارجي وترتيب النتائج والبحث في قواعد البيانات وغيرها؛ إلا أن الباحثين عن المعلومات لا يكادون يرون ذلك الفارق في عمليات البحث؛ حيث يلجأ المستخدم إلى أداة بحث بغض النظر عن ماهية هذه الأداة، وذلك إنما يرجع إلى أسباب عدة منها(١) :-

١. أن الاختلافات في تنظيم المعرفة ومصادر المعلومات الإلكترونية من صفحات ومواقع الويب هو أمر لا يتم أمام مستخدمي الإنترنت.
٢. أن الكثير من أدلة البحث الجيدة مثل Yahoo؛ لا يستخدم بكثرة من قبل الباحثين، نظرًا لعدم تفحص المستخدم لمثل هذه الأدلة.
٣. أن محتوى الويب من مصادر المعلومات يتزايد بمعدل مرتفع، مما يجعل الباحث يلجأ لأدوات بحث مختلفة.
٤. أن كل من محركات البحث وأدلة البحث يكمل بعضها الآخر في عملية البحث العلمي.
٥. أن الاختلافات بين محركات البحث والأدلة البحثية إنما هي اختلافات ضبابية لا تضح معالمها طالما وجد بحث ونتائج يمكن الخروج بها.

ولقد كان نتيجة صغر الفروقات بين خصائص محركات البحث والأدلة البحثية أن اتجهت بعض المواقع إلى الجمع بين النوعين أي الجمع بين صفة محركات البحث وبين الخصائص المميزة لأدلة البحث وهو ما سيطلق عليه في هذه الدراسة مسمى "أدوات البحث". وفيما يلي سيتم تناول أدوات البحث بالتفصيل.

٢/١ أدوات البحث على الويب

إن الأهمية التي تأخذها هذه الأدوات تأتى من الدور الذي تؤديه في الويب باعتبارها جسرًا يهدف إلى الوصول إلى المعلومات المحددة؛ حيث تعمل هذه الأدوات عمل اختصاصي المعلومات في اقتناء وسائط المعلومات في مختلف المجالات ثم تنظيمها وإعدادها الإعداد الفني لوضعها قيد الاسترجاع في أي وقت

---

(١) ibid.

لمختلف الباحثين وكل حسب حاجته الموضوعية. ولولا هذه الأدوات لأصبحت الويب مستودعا يحمل بداخله معلومات العالم أجمع في شكل غير منظم وعشوائي لا يوفر أي وسيلة للبحث سوى مسار أو عنوان الموقع URL، مما ينعكس على المستفيد كالذي يبحث عن ضالة في فلاة.

وبذلك يمكن تحديد أهمية وجود أدوات بحث الويب في سببين هما:-

١. أن كم المعلومات على الإنترنت ينمو بالسرعة والحجم الذي لا يمكن لمستخدم ما معرفة ما هي المواقع التي تهمه وأين توجد؟
٢. أن الأغلبية من المستخدمين للويب حتى الماهرين منهم لا يستطيعون الوصول في عمليات البحث إلى مستوى الدقة المناسب والتحكم في عمليات الاستدعاء، ذلك لأنه يبحث عن شيء غير محدد أمامه.

١/٢/١ كيفية عمل أدوات البحث

تعمل أدوات البحث في الويب بشكل يشبه تمامًا طريقة العمل في مكتبة تقليدية، من حيث الإضافة والتزويد، والإعداد الفني والتنظيم، والبحث والاسترجاع، ويمكن تمثيل تلك العمليات كما يلي :-

أولًا: الإضافة والتزويد

تختلف الأدلة البحثية عن محركات البحث في الإضافة والتزويد، فالأولى تعتمد على القوى البشرية، والثانية تعتمد على البرامج الآلية. فالأدلة البحثية تستخدم الأفراد في مرحلة الإضافة والتزويد عن طريق تصفح الإنترنت ثم تجميع المواقع التي وقع عليها الاختيار، وفهرستها وتكشيفها لتوضع ضمن قاعدة بيانات الدليل، أما محركات البحث فتعتمد على برنامج الزاحف أو العنكبوت. وهذا يعني أن حجم التغطية في الأدلة البحثية ليست بالقدر الكافي لمواقع الويب المتنامية العدد. فالأدلة تعمل من منظور الأعداد القليلة التي يمكن التحكم فيها وإعدادها بما يحقق للباحث أكبر قدر من الدقة والتحقيق.. بينما تعمل محركات البحث على تغطية واسعة للمواقع على الويب.

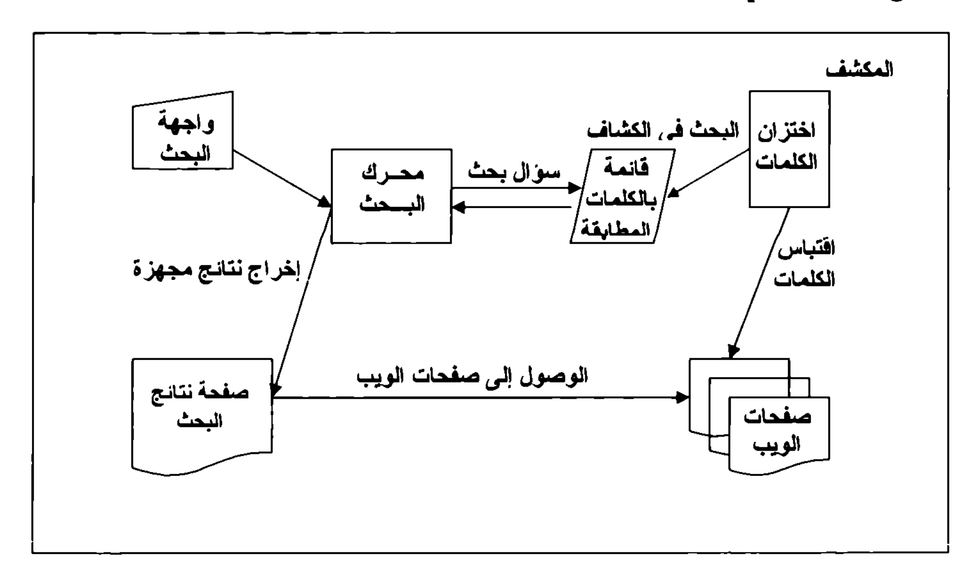
ومن ناحية أخرى، فإن التزويد والإضافة للأدلة البحثية يكون عن طريق الطلبات التي تقدم من أصحاب المواقع ذاتهم سواء كانت تجارية أو حكومية أو علمية أو شخصية أو غير ذلك من الأنواع الأخرى. أما الإضافة والتزويد في

محركات البحث فيكونان عن طريق الزاحف الذي يتحرك بين روابط الصفحات المختلفة ومن ثم تحليل كل الروابط داخل صفحات الموقع المجدد ثم زيارة هذه الروابط؛ إلا أن برامج الزاحف تختلف فيما بينها من حيث مستويات تحليل صفحات المواقع؛ حيث تكتفي بعض البرامج بالمستويين الأول والثاني لصفحات موقع الويب، وتتعدى بعض برامج الزاحف ذلك إلى المستويات الثالث والرابع. وهو ما يؤثر بشكل أساس على أعداد صفحات الإضافة اليومية لبرنامج الزاحف. وهذا يترتب عليه فترة التحديث التي قد تتأخر بعض الشيء. فالتزويد في محركات البحث يتم في فترات متقاربة كل ساعة أو كل يوم، وقد يكون في فترات متباعدة أسبوعيًا أو شهريًا؛ إلا أنه يمكن اعتبار هذه الفترات دورة يتم فيها الزاحف العمل إلى العودة مرة أخرى لمثل هذه المواقع لتبيان ما بها من تحديث، أو تغيرات في متن هذه الصفحات.

ثانيًا: الإعداد الفني والتنظيم

بعد أن تنتهي المرحلة الأولى بإضافة نسخة من موقع الويب إلى قاعدة بيانات محرك البحث، يبدأ عمل الجزء أو البرنامج الثاني في أدوات البحث الذي يعرف بالمكشف أو المفهرس، حيث يقوم بإعداد تسجيلات ببليوجرافية لمواقع الويب حديثة الإضافة إلى قاعدة بيانات أداة البحث، وتعمل هذه التسجيلة وفقاً لما ورد في المكتبات والمعلومات عن (التمثيل المكثف) حيث لا يتم اختزان موقع الويب بأكمله في عدد من أدوات البحث، وإنما حقول وبيانات منها مسار هذا الموقع للربط به والوصول إليه، وكل ذلك في شكل تسجيلة كما هو الحال في الفهارس الآلية للمكتبات OPAC. وتتم عملية الفهرسة في حالة الأدلة البحثية بشكل يدوي. أما في حالة محركات البحث فإنها تتم آليا، وفي حالة التكشيف والفهرسة الآلية فإن برامج التكشيف الآلي تستخدم بنية صفحة الويب في التعرف إليها ووضع كلمات أو مصطلحات معبرة عن الموضوع أو الموضوعات التي تضمها صفحة الويب؛ حيث يفحص البرنامج هنا كل أجزاء الصفحة التي تحتوي على كلمات تخص الموضوع من كلمات العنوان، وكلمات النص، والكلمات المفتاحية التي وضعها مصمم صفحة الويب في رموز meta tags وتجمع هذه الرموز كل الكلمات التي يرى مصمم الصفحة أنها تعبر عن موضوع أو موضوعات صفحة الويب. وبعد تفحص كل الجوانب السابقة يتم وضع كلمات مفتاحية داخل تسجيلة هذا الموقع أو وضع كل كلمات

النص الطبيعية في وضع نشط ليتم مضاهاتها بعد ذلك مع كلمات البحث التي يضعها الباحث عند إجراء البحث. وتختلف مواضع التكشيف من برنامج إلى آخر؛ حيث تعمل بعض برامج التكشيف على الاكتفاء بالجزء الأول فقط من بنية اللغة المعيارية لصفحة الويب، كما تعمل بعض البرامج على تكشيف عناوين الصفحات بجانب مسار الصفحة، وتمتد قدرة برامج التكشيف في أحيان أخرى إلى تكشيف كامل متن صفحة الويب وهو ما يجعل محركات البحث في هذه الحالة تمتاز بوفرة في الاستدعاء مع انخفاض نسبة التحقيق. ويبين الشكل رقم (١-١) طريقة عمل برنامج المكشف في دائرة عمل محرك البحث:



الشكل رقم (١-١) يوضح عمل برنامج المكشف في دائرة عمل محرك البحث

ثالثا البحث والاسترجاع

يعد الاستدعاء من أهم مقاييس كفاية النظام. لكن الأمر لا يقتصر على النظر إلى النتائج، بل يتعدى إلى ماهية التنظيم داخل نظام الاسترجاع؛ فمحركات البحث تعتمد على البحث الحر في كامل نص الموقع مع إمكانية البحث في العنوان والرموز المعيارية (metatags) وبذلك فإن الاستدعاء سوف يكون في أعلى درجاته.. بينما يكون التحقيق في أقل حالاته، ذلك لأن آلية التكشيف التي تعتمد عليها محركات البحث هي استخدام اللغة الطبيعية وهذا إنما يعكس بدرجة عالية أن كل

كلمات النص سوف تصبح كلمات مفتاحية يمكن من خلالها استدعاء الوثيقة (موقع الويب) اعتمادًا على مضاهاة الكلمات البحثية. أما الأدلة البحثية فإنها تعتمد في البحث على كلمات كشفية ومصطلحات يتم وضعها بالتدخل البشري تعبر بصدق عن الموضوع. وبالتالي فإنه يستبعد أية وثائق أو مواقع غير ذات صلة، ويولد تحقيقًا عاليًا مع استدعاء منخفض .

ومن ناحية أخرى تكون النتائج في محركات البحث مرتبة حسب أهمية الموقع للموضوع. ويختلف الوزن النسبي لموقع الويب وتقدير أهميته للموضوع إما اعتمادًا على عدد تكرار الكلمات المفتاحية التي ذكرت في السؤال، أو أماكن تواجد الكلمات المفتاحية والموقع بحيث إذا ما وردت في العنوان يكون للموقع أهمية أكثر من وجودها في النص نفسه. وبحسب الأهمية يأتي ترتيب النتائج بالموقع فالأكثر أهمية للموضوع يأخذ الرقم (١) في الترتيب ثم ما يليه من المواقع على حسب هذا الوزن ... وهكذا. أما بالنسبة للأدلة البحثية، فإنه عادة ما تحدد هذه الأهمية من قبل المفهرس نفسه الذي يحدد موضع الموقع في الترتيب.

وفي كلتا الحالتين، فإن البحث والاسترجاع يعتمدان على إستراتيجيات تستخدم معاملات المنطق البوليني، وآلية بحث اللغة الطبيعية، والبحث بالتماثل، والبحث بالمفهوم، والبحث بالتقارب بين الكلمات near, next، والبحث بالمتضادات، والبحث بالترجمة... وغيرها الكثير من معاملات وآليات بحث تعمل جميعا لأداء أفضل تحقيق للباحث عن المعلومات وسيأتي ذكرها بالتفصيل.

٢/٢/١ إستراتيجيات البحث في أدوات البحث

إن من أهم المصطلحات التي تذكر بذكر محركات البحث هو مصطلح إستراتيجيات البحث. ويمثل هذا المصطلح أهمية كبيرة في فهم محركات البحث ذلك؛ لأنها تمثل أحد الأعمدة الثلاثة التي يقوم عليها محرك البحث؛ الإضافة والتكشيف والبحث. وقد ذكر Paul Pedley أن إستراتيجيات البحث تقوم على خمسة عناصر أساسية هي[١] :-

١. أن يتم صياغة أسئلة البحث الصياغة التي تحدد الهدف من طرح السؤال.

---

(١) ibid.

٢. أن تحدد المفاهيم المهمة من كل كلمات السؤال المطروح.

٣. أن تحدد مصطلحات أو كلمات البحث التي تصف تلك المفاهيم.

٤. أن تحدد المترادفات المختلفة التي تعبر عن نفس الموضوع والمفاهيم المطلوبة.

٥. أن ينشأ منطق بحث يمكن استخدامه من جانب الباحثين.

هذا عن الأساس الذي تسير عليه إستراتيجيات البحث؛ في أدوات البحث، إلا أن التطبيق والتنوع في إستراتيجيات البحث هو الواقع القائم بالفعل داخل تلك الأدوات. ونظرًا لأن جميع محركات البحث تسعى للزيارة الدائمة من جانب المستخدمين، فإن هذه المحركات لا تدخر جهدًا في تحسين إستراتيجياتها وتقوية نظم الاسترجاع لديها. وتضم إستراتيجيات البحث أنواعا عدة. ولقد ضمت دراسة Anne Clyde مختلف إستراتيجيات البحث المستخدمة داخل محركات البحث. ويمكن تناول هذه الإستراتيجيات كما يلي[١] :-

١. النوع الأول: إستراتيجيات بحث اللغة الطبيعية وهنا تتيح محركات البحث استخدام لغة الحوار الطبيعي من جانب المستخدم، حيث لا يتوقف الأمر على إدخال كلمات مفتاحيه، وتعمل بعض المحركات مثل AskJeeves على أن تناسب فئات الأطفال بجانب كبار السن. وعلى ذلك فإن محركات البحث لا تألو جهدًا في جعل البحث أكثر يسرًا على الباحث وفي الوقت ذاته تجعل التحقيق أكثر دقة في النتائج.

٢. النوع الثاني : معاملات المنطق البوليني Boolean Operators وتعمل هذه المعاملات على ربط الكلمات المفتاحية معًا مكونة حدودًا موضوعية ترسم شكل الموضوع النهائي. وتستخدم المعاملات؛ AND للجمع، NOT للاستثناء، OR للبدائل.

٣. النوع الثالث : استخدام البحث الحر : ويشتمل هذا النوع على ما يجعل الباحث يتفادى الخطأ في الإملاء، مما يؤثر بدوره على مصطلحات البحث، كما أنه

---

(١) Clyde, Anne. 2000. Search engines: an overview, teacher librarian, Visited at: 18/9/2002,: http://www.hi.is/~anne/websearch_bibliography.html

يجعل محركات البحث تستخدم كل الأشكال المختلفة للكلمة الواحدة مثل إدارة تسترجع إدارة، إداري، إداريون، إدارات... الخ من الأشكال الأخرى.

٤. النوع الرابع : إستراتيجيات بحث المفرد والجمع : وهذا النوع من إستراتيجيات البحث يعمل على استرجاع المفردات بأشكالها المختلفة من مفرد وجمع، وعلى ذلك فإن كلمة "طفل" تسترجع "أطفال"، وكلمة مكتبة تسترجع مكتبات.

٥. النوع الخامس : إستراتيجية البحث بالقرب. وهذا النوع له معاملاته الخاصة في صياغة إستراتيجية البحث مثل مقترن بـadjacent، قريب إلىnear، متبوع بـ allowed by ويتحكم بذلك الباحث في الجمع بين مصطلحين من مصطلحات البحث ليخرج الموضوع على الشكل الذي يرتضيه.

٦. النوع السادس : إستراتيجية البحث بحساسية الحالة : التي يتم فيها البحث عن معنى المصطلح إذا وجد على شكل معين، فتكون Earth هي الكوكب أما earth فهي التربة. ومن المحركات التي تتيح هذا النوع. ومثال آخر كالبحث بالأسماء العربية ذات الهجائية المختلفة " هبه - هبة، على - علي، أحمد - احمد "

٧. النوع السابع : إستراتيجية البحث بالعبارات : ويعمل هذا النوع على جمع مصطلحين أو أكثر للتعبير عن الموضوع، ولا تصل العبارات هنا إلى جمل أو أسئلة. والعبارات البحثية تكون مثل " النظام الشمسي، مكتبات المدارس " .

٨. النوع الثامن : إستراتيجية البحث بالحقول ويتم البحث في حقول التسجيلة الواصفة لوسيط المعلومات: موقع أو صفحة الويب، كأن يتم البحث في حقل العنوان مثل t:=Islam حيث :t هي حقل العنوان title. أيضا "عن" للعنوان، "مو" للمؤلف.

٩. النوع التاسع : إستراتيجية النتائج الأولية، ويقوم هذا النوع من البحث على عرض النتائج الأولية للبحث، ثم يحدد الباحث العناوين الأكثر قربا للموضوع، ويأتي دور محرك البحث في استرجاع كل النتائج المشابهة لما اختاره الباحث وقد أطلقت عليها (more like this Clyde).

ويرى الباحث أنه يجب أولا الوقوف على الفارق بين مصطلحات إستراتيجية البحث وآلية البحث، ذلك لما بينهما من تداخل ؛ حيث يمكن النظر إلى إستراتيجية

البحث على أنها الشكل النهائي لصياغة كلمات البحث والربط بينها بآليات البحث، وهي بمثابة الحصيلة النهائية من الكلمات وأدوات الربط التي تدخل إلى قاعدة بيانات نظام الاسترجاع. أما آليات البحث فهي الأشكال المختلفة من أساليب بحث نظم الاسترجاع؛ حيث تتيح بعضها استخدام آلية المنطق البوليني بمختلف معاملاته، ويتيح البعض الآخر استخدام الحقول أو استخدام آلية البحث بالبتر الضمني بمعاملاته المختلفة. وعلى ذلك فإن الأنواع السابقة من إستراتيجية البحث التي ذكرت في دراسة Clyde قد شملت بداخلها نماذج عدة من آليات البحث مدخلة عليها مفاهيم إستراتيجيات البحث.

وتقسم عناصر البحث داخل نظم الاسترجاع من الأعم إلى الأخص كما يلي؛ أولًا: إستراتيجيات البحث وتضم الشكل النهائي المدخل إلى قاعدة البيانات، ثانيًا: آليات البحث وهي أساليب لبحث وحدات قواعد البيانات مثل المنطق البوليني والبتر الضمني والبحث بالحقول، ثالثًا: معاملات البحث وهي العناصر المكونة لآليات البحث مثل معاملات and, or, not لتكوين الية المنطق البوليني ومعاملات ?, * آلية البتر الضمني ومعاملات ti, su, au المكونة لآلية البحث بالحقول.

وقد اختلطت الأنواع التي ذكرتها Clyde من إستراتيجيات البحث فضمت بينها أنواعًا من آليات البحث والتي تختلف بدورها عن مفهوم إستراتيجية البحث، ذات المفهوم الأعم عن الآليات؛ فإذا ما أخذ البحث بالحقول مسمى إستراتيجية البحث، يأخذ المنطق البوليني والربط مختلف العلاقات بين الحقول مسمى آلية البحث. ومن ثم فإن مستويات بحث نظم استرجاع المعلومات يحتاج إلى نظم بحث تتكون من إستراتيجيات بحث مثل البحث بالكلمات الحرة أو الكلمات المقيدة، وتحتوي إستراتيجيات البحث بدورها على آليات بحث مثل المنطق البوليني أو البحث بالتقارب، ولا تقوم هذه الآليات بعملها إلا من خلال معاملات البحث مثل Or, And داخل المنطق البوليني.

إضافة إلى ما سبق يمكن إضافة أنواع أخرى من إستراتيجيات البحث مثل البحث بالمفهوم، وهو ما يطلق عليه Concept-based searching، ولقد برز هذا النوع نتيجة لما يواجهه البحث بالكلمات المفتاحية من صعوبات وأهمها الاشتراك اللفظي بين الكلمات في حالة البحث إذا ما استخدمت كلمة مفتاحيه واحدة في البحث؛ فإن استخدام كلمة واحدة في البحث مثل " قلب " في اللغة العربية أو "Heart" في

اللغة الإنجليزية للبحث عن موضوع القلب من الناحية الطبية، ينتج عنه الكثير من النتائج التي تبتعد عن الموضوع مثل الإصابات التي تتجه صوب كلمة القلب في الرومانسية.

وتعمل الكثير من محركات البحث الآن على الإنترنت مثل Excite على استخدام البحث بالمفهوم. ولعله من نافلة القول أن استخدام المفهوم لا يكون فقط في مرحلة البحث وإنما في مرحلة التكشيف بالتعيين، الذي يعتمد في مجمله على كلمات تعبر عن مفاهيم يتم استخدامها لتمثيل الوثيقة أو مصدر المعلومات. وقد ذكرت Linda Barlow أن البحث بالمفهوم يختلف عن البحث بالكلمات في الاهتمام بما تعبر عنه الوثيقة Aboutness وليس بما تحويه الوثيقة من كلمات واسترجاعها باستخدام منطق المضاهاة. وتسترسل Barlow في الحديث عن عمل البحث بالمفاهيم بأن ذلك يأتي بالاعتماد على وجود مجموعة من الكلمات التي تعبر مجتمعة عن موضوع واحد أو جانب من موضوع عام وذلك كما في المثال التالي، فإذا ما وجدت كلمة "heart" بجانب كلمات أخرى مثل "blood, cholesterol, attack, lung and stroke" فإن هذا يدل على تناول القلب من الناحية الطبية، أو الموضوع الفرعي القلب من مجال الطب. أما في حالة أن تأتي كلمة "heart" وسط مجموعة من الكلمات مثل "love, valentine, flowers, candy" فإن ذلك يعني أن الوثيقة تتناول القلب في إطار الحديث عن الرومانسية. مما يترتب على كل حالة من الحالتين السابقتين إبعاد مجموعات المصادر التي لا ترتبط بالموضوع المراد البحث عنه. وقد أوردت Barlow في دراستها أن مثل هذه النظم التي تعمل بالمفهوم لا تتعدى كفايتها إلى الناحية العملية، إنما هي فقط نظم مثالية من الناحية النظرية.[١]

١/٣ محركات البحث

إن الكم المعلوماتي الذي تحمله الويب لا يمكن الاستفادة منه؛ إلا عن طريق التنظيم والإعداد. وتعد محركات البحث مرشد الباحث ودليله في الوصول إلى المعلومات على الويب. وتعتمد محركات البحث على برامج وإن تبدو أمام الباحثين سهلة؛ إلا أنها دقيقة في أدائها، ذلك لأنها تجمع المعلومات من مواقعها ثم تنظمها

---

(١) Barlow, Linda. A Helpful Guide to Web Search Engines, Monash Informat:ion Services, 2002, 11/112002, Visited at: http://www.monash.com/spidap.html

بعمليات الفهرسة والتكشيف وتتيح بعد ذلك آليات للبحث داخل قواعد بياناتها لتعطي بعد ذلك ثمرة هذا الأداء في شكل مواقع تناسب مصطلحات البحث وتعد الويب بالفعل مثل الصندوق المغلق الذي لا يعرف أحد ما بداخله، كما جسد ذلك Paul Pedley في الحديث السابق عن الويب غير المرئية والتي لا يضيء ظلمة بحث المعلومات داخلها سوى محركات وأدلة البحث.

١/٣/١ مفهوم محرك البحث

تعددت التعريفات لمحركات البحث، ومن التعريفات التي وضعت لمحركات البحث: أنها "عبارة عن أداة تقوم بالبحث في و/ أو عن مصادر المعلومات على الإنترنت ـ والمصادر هنا يقصد بها المعلومات على المواقع وتخزين عناوينها على مرصد البيانات الخاص بها، ثم تقوم بإتاحتها للمستفيدين كل حسب المصطلح أو المصطلحات المستخدمة في البحث ومن ثم تمكن المستفيد من الوصول إلى مصادر المعلومات المختلفة على الإنترنت. ويتم تجميع هذه المصادر إما بطريقة الية spidering or crawling أو بطريقة بشرية بواسطة الإنسان."[١]

وقد عرف زين عبد الهادي محركات البحث في البيئة العربية على أنها "أدوات بحث تعمل من خلال إستراتيجيات بحث محددة المنطق البوليني أو إستراتيجيات بحث مفتوحة باللغة العربية، وذلك للبحث في حقول أو وثائق نصية، والأكثر من ذلك أنها تبحث عن أشياء كالمادة المصورة والخرائط والأشكال الأخرى في بيئة محددة هي شبكة الإنترنت وذلك يعني أنها تبحث في ملايين المواقع ومليارات الكلمات في وقت محدد وتتميز بسرعة الاستجابة وعادة ما تكون إجاباتها إما مواقع على الإنترنت تتوافر فيها كل المصطلحات التي تم البحث عنها أو بعضها، أو مواقع محددة سلفا من خلال ما يعرف بأدلة البحث"[٢].

لقد اعتمدت التعريفات السابقة لكل من خالد محمد رياض وزين عبد الهادي على التفصيلات المختلفة التي يتعرض لها محرك البحث في أداء مهامه

---

(١) خالد محمد رياض. **أدلة ومحركات بحث شبكة الإنترنت: دراسة مقارنة** -. الاستراتيجية العربية الموحدة للمعلومات في عصر الإنترنت ودراسات أخرى، الاتحاد العربي للمكتبات والمعلومات.- تونس: (أكتوبر١٩٩٩). ص ١٣٩.

(٢) زين عبد الهادي. محركات البحث على شبكة الإنترنت: دراسة تجريبية مقارنة، **مجلة المكتبات والمعلومات العربية**، ع ٢ أبريل ٢٠٠٢.ص ١٠.

على الويب، وعلى ذلك فقد اتخذت التعريفات السابقة الشكل الإجرائي للحديث عن محرك البحث.

و قد ذكرت بعض مواقع الويب تعريفات عن محركات البحث منها ؛ " إن محرك البحث هو بصفة عامة أداة تمكن مستخدمي الويب من البحث بمصادرها، وأنه يمكن أن يعتمد على البرامج أو الإنسان في عمله"(١) وهذا التعريف قد عمل على تلخيص مهمة محرك البحث وإبرازها من ناحية مستخدمي الويب دون التطرق إلى الإجراءات التي يقوم بها في تنفيذ هذه المهمة. وثمة تعريف آخر هو "أنه عبارة عن أداة لإيجاد المعلومات على الإنترنت، ويتكون من مكونات خمسة هي أولا العنكبوت أو الروبوت، ثانيا المكشف، ثالثا قاعدة البيانات، رابعًا برامج البحث، خامسًا واجهة الويب"(٢). وقد اختلف هذا التعريف عما سبقه من ناحية التركيز على مكونات محرك البحث، وإن كان الباحث يرى أنها قد تقتصر فقط على ثلاثة مكونات هي الزاحف والمكشف والبحث.

وعلى ما سبق من تعريفات يرى الباحث أن أفضل تعريف لمحركات البحث هو " إحدى أدوات بحث استرجاع المعلومات على الويب، وتعتمد على ثلاثة برامج أساسية في ذلك هي الزاحف والمكشف وآليات البحث، بما يوفر القدرة على الإضافة والبحث. "

والجدير بالذكر أن هناك فرقًا بين مصطلحي محركات البحث وآليات البحث، فالأول إنما يعني برامج لبحث مصادر معلومات الويب، أما الثاني فهو طرق البحث داخل قواعد بيانات محركات البحث مثل آليات المنطق البوليني، والبتر الضمني، واللغة الطبيعية، والبحث بالمعنى أو المفهوم... الخ. وعلى ذلك فإن ثمة وجهين تعمل بهما محركات البحث؛ الوجه الأول وهو الذي تتعامل به مع الإنترنت في عملية التجميع والفهرسة والتكشيف لمواقع الويب والوجه الثاني الذي تتعامل به مع مستخدمي الويب طلبًا للمعلومات. والوجه الثاني قد يكون مشرقًا أمام بعض

---

(١) Sullivan, Danny. Search Engine Optimization & Marketing Glossary, Submitawebsite, Inc, 2004, Visited at 12/6/2004, http://www.submitawebsite.com/seo_glossary.html

(٢) What is a Search Engine? Metamend.inc, 2004, Visited at 12/6/2004, http://www.metamend.com/search-engine-map.html

الباحثين بالحصول على المعلومات التي يريدونها بسهولة ويسر، وقد يكون عبوسًا أمام البعض الآخر لما يلاقيه هؤلاء من نتائج غير مطابقة لموضوعاتهم وأسئلتهم البحثية، تنتج عن استخدام المصطلحات غير المناسبة لاستدعاء مصادر الموضوع.

فالأمر هنا ليس مجرد الدراية بآليات البحث والتمرس بها، بل يتعداه إلى حد وعي الباحث بجوانب الموضوع؛ فعلى سبيل المثال فإن سؤالا مثل (ما هو التدخين ؟) فإنه ينطوي على جانب موضوعي واحد هو (التدخين)، وسؤالا عن (هل يسبب التدخين السرطان ؟) فإنه ينطوي على جانبين هما (التدخين، والسرطان)، كما إن سؤالا عن (هل يعلم الأطفال أن التدخين يسبب السرطان ؟) يتشابك فيه مجموعة الجوانب " الأطفال والتدخين والسرطان". وعلى ذلك فلا بد أن يكون الباحث على دراية بماهية الموضوع وجوانبه المختلفة لكي يستطيع تحليله ووضع المصطلحات التي تقيه أخطاء وتشويش النتائج في محركات البحث التي لا تعرف سوى مطابقة أو مضاهاة كلمات النص جميعها مع مصطلحات البحث.

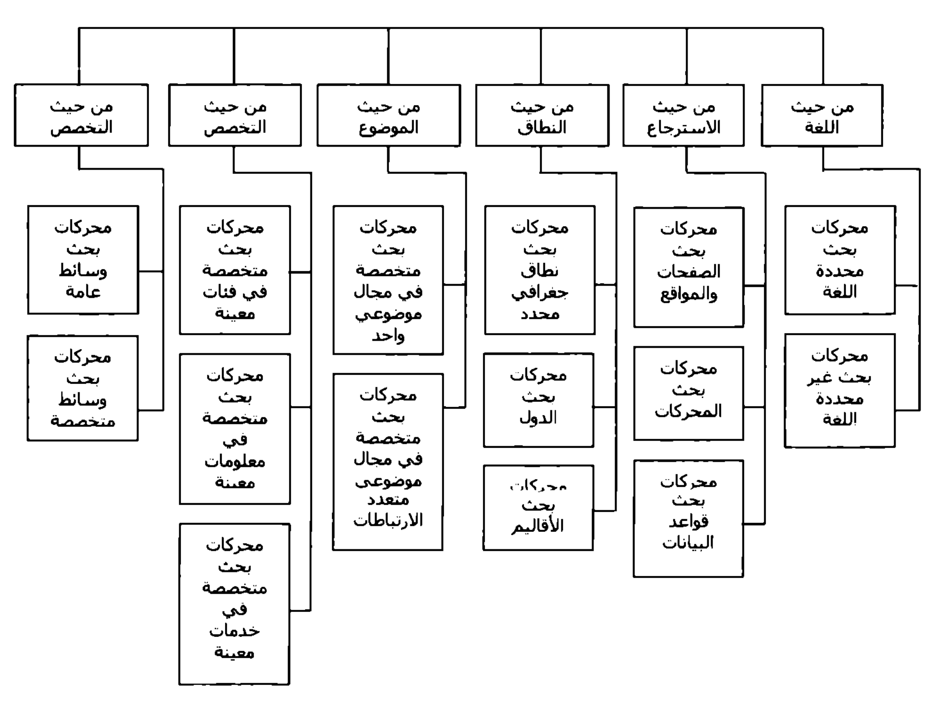
ومن ثم فيمكن أن نلخص الأسباب التي تدفع الباحثين لاستخدام محركات البحث في:-

١. أن يكون لدى الباحث فكرة أو موضوع مخصص داخل ذهن الباحث.
٢. أن يهدف الباحث الوصول إلى موقع محدد أو مخصص.
٣. أن يهدف الباحث بحث النص الكامل لملايين من صفحات ومواقع الويب.
٤. أن يهدف الباحث استدعاء كم كبير من مواقع وصفحات الويب.
٥. أن يهدف الباحث بحث نوع معين من صفحات أو ملفات الويب أو مواقع للغة معينة.
٦. أن يرغب الباحث في الحصول على ميزات تدعم بحث الويب مثل وزن النتائج المسترجعة أو البحث بالمفاهيم.

١/٣/٢ أنواع محركات البحث

تعتمد محركات البحث على نوع واحد من البرامج المستخدمة التي يتشابه أداؤها في كل أنواع المحركات وسوف يكون التقسيم هنا ليس لإبراز هذه الأنواع وإنما لبيان الجوانب التي يمكن أن تتناول أنواع محركات البحث عن طريقها.

ويوضح الشكل رقم (١-٢) أنواع محركات البحث مقسمة وفق التخصص.وفيما يلي شرح لهذه الأنواع:-



الشكل رقم (١-٢) يوضح أنواع محركات بحث الويب

١/٢/٣/١ أنواع المحركات من حيث اللغة

تستخدم الغالبية العظمى من المحركات اللغة الإنجليزية بالإضافة إلى بعض اللغات الطبيعية الأخرى في البحث عن الصفحات والمواقع المتاحة على الويب؛ إلا أن هناك عددًا من المحركات تسمح باختيار لغة النص الذي يبحث فيه، واستبعاد ما دونه من اللغات. ويمكن بذلك تقسيم محركات البحث حسب لغات النص التي تبحث فيها إلى ما يلي:-

١- **محركات بحث غير محددة اللغة**: حيث لا يعطي محرك البحث هنا خيارًا للغة النص، وإنما يقوم محرك البحث هنا بمضاهاة الكلمات التي جاءت في إستراتيجية البحث في أي صفحات وجدت. ويفيد مثل هذا النوع من المحركات الباحثين الذين لا يسعون إلى تحقيق عالٍ وإنما يأتي الاستدعاء في

المقام الأول. هذا بالإضافة إلى أن لغات النشر الإلكتروني لم تعد تقتصر على مجموعة لغات معينة وإنما أصبح كل من يمتلك حاسبًا آليًّا ومكانًا على أحد خادمات الإنترنت باستطاعته النشر وإلقاء المعلومات التي يريدها في فضاء الويب. ومن ثم أصبح من غير الممكن أن يلم الباحث بكل اللغات التي ينشر بها الإنتاج الفكري في موضوع بحثه. ومن نماذج محركات البحث غير محددة اللغة محرك البحث Altavista[١] ومحرك البحث Google[٢] ومحرك البحث Alltheweb[٣].

٢- **محركات بحث محددة اللغة**: يأتي هذا النوع عكس سابقه، حيث يسمح المحرك بالتحديد في البحث بلغات النصوص المنشورة بها صفحات الويب بطرق مختلفة منها؛ الطريقة الأولى : أن يقوم محرك البحث باستدعاء كل الصفحات التي تحوي الكلمات المفتاحية ضمن نصوصها بغض النظر عن اللغة، وإذا ما أراد الباحث تحديد لغة النص، فإن محرك البحث يمده بقائمة من اللغات الطبيعية المختلفة يختار منها الباحث ما يناسبه فيتم استبعاد الصفحات التي لا تقع مفردات نصها ضمن اللغة المختارة، وهنا تدخل اللغة كعامل محدد إضافي لنوع النتائج والوحدات التي يخرجها محرك البحث. الطريقة الثانية : أن يفرض محرك البحث مسبقًا لغة النص التي يرغب البحث عنها، وهذه ترجع إلى الحالات التي يكون فيها محرك البحث مختصًا بمنطقة محددة تعتمد لها لغة خاصة بها.

ومن نماذج محركات البحث محددة اللغة:

- محرك البحث Google[٤]
- محرك البحث باللغة الإسبانية [٥]
- محرك البحث باللغة الفرنسية [٦]

---

(١) http://www.Altavista.com
(٢) http://www.Google.com
(٣) http://www.Alltheweb.com
(٤) http://www.Google.com
(٥) http://www.sportec.com/
(٦) http://www.pratique.fr/

٢/٢/٣/١ أنواع المحركات من حيث الاسترجاع

يقصد بأساليب الاسترجاع هنا هو كيف يتم استرجاع مواقع وصفحات الويب هل بالأسلوب المباشر أي باللجوء إلى بيانات الفهرسة للصفحات والمواقع، أم بالأسلوب غير المباشر أي من خلال أدوات بحث أخرى. وتنقسم محركات البحث من حيث أساليب الاسترجاع إلى ثلاثة أنواع:

١- **محركات بحث الصفحات والمواقع** : وهي الأكثر انتشارًا وتعد بمثابة محركات بحث من الدرجة الأولى وتعمل على جمع وتنظيم وبحث مواقع وصفحات الويب بشكل مباشر من خلال الاطلاع على بيانات الفهرسة الخاصة بها، مثل URL الخاص بالصفحة مع مجموعة الواصفات التي تحدد الهوية الموضوعية لهذا الموقع. أو من خلال التعرف إلى كلمات وردت في النص.

ومن نماذج محركات بحث الصفحات والمواقع:

محرك البحث Google[١]

محرك البحث Yahoo[٢]

محرك البحث Altavista[٣]

٢- **محركات بحث المحركات** : ويعرف هذا النوع بالإنجليزية باسم Meta Search Engines. وهي تعد بمثابة محركات من الدرجة الثانية، وهي تبحث في نتائج البحث التي تأتي بها محركات بحث الدرجة الأولى. وهذا النوع يشبه مجازًا ما تكون عليه ببليوجرافيات الببليوجرافيات؛ إلا أن بعض محركات بحث المحركات تتعدى الحصر لنتائج محركات الدرجة الأولى إلى البحث والتنقيب المباشر في صفحات الويب. وتعرض هذه المحركات نتائج البحث إما بعرض نتائج البحث حسب كل محرك أو أن يتم عرضها بصفة عامة دون التحديد.

ومن نماذج محركات بحث المحركات

محرك البحث Alltheweb[٤]

---

(١) http://www.Google.com

(٢) http://www.yahoo.com

(٣) http://www.mltavista.com

(٤) http://www.Alltheweb.com

محرك البحث MetaCrawler[١]

محرك البحث Mamma[٢]

٣- **محركات بحث قواعد البيانات :** يعمل هذا النوع على اكتشاف الجزء غير المرئي من الويب، حيث مازال يطلق على قواعد البيانات مصطلح the invisible web. وتحاول هذه المحركات إخضاع محتويات قواعد البيانات من المقالات ومصادر المعلومات للبحث أمام الباحثين دون اللجوء إلى مواقع تلك القواعد، ذلك من أجل إتاحة كل مصادر معلومات الويب لإفادة الباحثين. ولقد وقفت طبيعة بناء قواعد البيانات حائلا أمام قدرة برامج الزاحف لاقتناء وإضافة الصفحات المتعمقة داخل هذه القواعد، كما تسعى محركات البحث على رفع قدرة برامج الإضافة بها لتحليل صفحات تلك القواعد بما يمكن معه اقتناء جميع وحدات المعلومات داخل قاعدة البيانات.

ومن نماذج محركات بحث قواعد البيانات

محرك البحث Invisibleweb[٣]

محرك البحث Webdata[٤]

٣/٢/٣/١ أنواع المحركات من حيث النطاق الجغرافي

يتمثل التخصص الجغرافي في أن يحصر المحرك بحثه في المواقع التي تقع أو تتبع حدودًا جغرافية محددة؛ حيث يقوم المحرك بجمع وتنظيم المواقع التي تخص منطقة أو إقليمًا جغرافيًا محددًا. وعادة يلجأ في هذه الحالة إلى البحث عن النطاق المعرف للدولة. وهذا النوع يوجد منه الكثير على الويب، ويمكن تقسيم محركات البحث من حيث التخصص الجغرافي إلى ما يلي:-

١- **محركات بحث ذات نطاق جغرافي محدد:** أي أن محركات بحث هذا النوع يمكن أن تفيد الباحثين بتقديم مواقع وصفحات الويب التي تنتمي إلى قارة من القارات، مثل المحركات التي تختص بمواقع القارة الأوربية أو تلك

---

(١) http://www.metaCrawler.com

(٢) http://www.mamma.com

(٣) http://www.Invisibleweb.com

(٤) http://www.Webdata.com

المحركات التي تهتم بمواقع وصفحات الويب التي تنتمي إلى قارة أمريكا الشمالية. وقد تتداخل هذه المحركات مع تلك الأنواع المتخصصة في اللغة وهي المناطق ذات اللغات المحدودة جغرافيًا.

ومن نماذج محركات البحث ذات النطاق الجغرافي المحدد

محرك بحث عن آسيا Lycos Asia[١]

محرك بحث عن كندا canadaspace[٢]

محرك بحث عن ولاية new Mexico[٣]

٢- **محركات بحث الدول** : تقتصر عملية الإضافة والبحث داخل هذا النوع فقط على مواقع وصفحات الويب المنتمية للدولة التي يعمل المحرك في نطاقها. وذلك من خلال محددات في تصميم مواقع الويب يمكن من خلالها الاستدلال على الدولة المنتجة للموقع. وقد يتعدى الأمر إلى المعنى الذي تقوم عليه الببليوجرافيا الوطنية: حيث تجمع ما ينتمي إلى الدولة بالمعنى الواسع مثل الذي تنشره الدولة، وما ينشر عنها، وما تنشره مؤسسات تابعة لها في أي موقع في العالم.

ومن نماذج محركات بحث المناطق والدول:

محرك البحث عن مصر Egyptsearch[٤]

محرك البحث عن اليابان SearchDesk[٥]

محرك البحث عن فرنسا Abondance[٦]

٣- **محركات بحث الأقاليم** : تعمل محركات بحث الأقاليم على توسيع نطاق بحث المواقع بشكل أكبر اتساعًا من محركات بحث الدول، وفي الوقت ذاته لا تكون على قدر الاتساع نفسه في المواقع الشاملة. ويأتي هذا النوع نتيجة لتشابه

---

(١) http://www.lycos.com

(٢) http://www.canadaspace.com

(٣) http://www.nmmagazine.com/cgi-bin/links.cgi

(٤) http://www.egyptsearch.com/search/

(٥) http://www.searchdesk.com/

(٦) http://webmasters.abondance.com/

مجموعة من الدول في اللغة والنواحي البحثية أو الاجتماعية. ويمكن أن يقال الثقافة في هذه الدول مثل الشرق الأوسط، أو الخليج العربي، أو دول شمال أوربا المعروفة باسم اسكندنافيا، أو دول جنوب شرق آسيا والصين وتايوان والهند وهونج كونج. وتوفر هذه المحركات الدرجة العالية من الاستدعاء لمواقع موضوع البحث مع وجود حدود يمكن الرجوع إليها للتحكم في النتائج. وفيما يلي أمثلة لهذا النوع

ومن نماذج محركات بحث الأقاليم

محرك البحث Almashriq (Middle East)[١]

محرك البحث Latin America[٢]

محرك البحث SearchEurope[٣]

١/٤/٢/٣/١ أنواع المحركات من حيث الموضوع

هناك الكثير من محركات البحث على الويب تقتصر في بحثها على تخصص موضوعي واحد مثل الطب، أو قد يتسع ذلك ليغطي عددًا من الموضوعات ذات الصلة كالطب والصحة والجراحة. وتأخذ هذه المحركات في الغالب شكل قواعد البيانات على الويب. وهذا النوع من قواعد البيانات يهتم بموضوع محدد ويجمع كل ما يمكن من مواقع ومصادر معلومات إلكترونية تفيد الباحثين والمهتمين بهذا الموضوع. ومن ثم فإن عمل محرك البحث هنا لا يتفق كثيرًا مع المعنى الذي يحمله مصطلح محرك البحث؛ إلا أن الباحث قد لاحظ ذكر أنواع قواعد البيانات في تصنيف محركات البحث داخل الفئة المتخصصة موضوعيا. وتنقسم محركات البحث من حيث الموضوع إلى نوعين رئيسين هما:-

١- **محركات بحث متخصصة في مجال موضوعي واحد** : وهي المحركات التي تهتم بالبحث في أحد الموضوعات دون النظر إلى بقية فروع المعرفة، ومن أهم الأمثلة على هذا النوع ما يلي:-

---

(١) http://almashriq.hiof.no

(٢) http://www.yupimsn.com/index.asp

(٣) http://www.searcheurope.com

ومن نماذج محركات البحث المتخصصة في موضوع واحد

محرك بحث متخصص في الطب Midisearch[1]

محرك بحث متخصص في القانون Lawcrawler[2]

محرك بحث متخصص في العلوم Search4Science[3]

٢- **محركات بحث متخصصة في مجال موضوعي متعدد الارتباطات**: وهي المحركات التي تعمل في مجال موضوعي يغطي مجالات معرفية ذات علاقة ببعضها البعض كالعلوم الاجتماعية.

ومن نماذج محركات البحث المتخصصة في موضوع متعدد الارتباطات:

محرك البحث في العلوم الصحية mwsearch[4]

محرك البحث في الإنسانيات eserver[5]

محرك البحث في إدارة الأعمال business[6]

١/٣/٢/٥ أنواع المحركات من حيث التخصص النوعي

وتتخصص هذه المحركات في تغطية صفحات ومواقع الويب التي تهم فئة معينة من الباحثين أو تغطي معلومات، أو خدمات معينة. وتنقسم محركات البحث هذه إلى ثلاثة أنواع رئيسة هي:-

١- **محركات بحث متخصصة في فئات معينة من المستفيدين** : ومن أمثلة هذا النوع محركات البحث المتخصصة في مواقع الأطفال.

ومن نماذج محركات بحث الأطفال

محرك البحث KidsClick[7]

---

(١) http://midisearch.net/

(٢) http://findlaw.com

(٣) http://www.search4science.com/

(٤) http://www.mwsearch.com/

(٥) http://www.eserver.org/

(٦) http://www.business.com/

(٧) http://www.kidsClick.com

محرك البحث Ask Jeeves for kids[١]

محرك البحث Yahooligans[٢]

٢- **محركات بحث متخصصة في معلومات معينة**: ومن أمثلة هذا النوع محركات البحث المتخصصة في الأخبار.

ومن نماذج محركات بحث الأخبار

محرك البحث Google News[٣]

محرك البحث Yahoo News[٤]

محرك البحث Altavista News[٥]

٣- **محركات بحث متخصصة في خدمات معينة**: ومن أمثلة هذا النوع محركات البحث المتخصصة في خدمات التسوق.

ومن نماذج محركات بحث التسوق:

محرك البحث Froogle[٦]

محرك البحث MSN Shopping[٧]

محرك البحث Yahoo Shopping[٨]

٦/٢/٣/١ أنواع المحركات من حيث التخصص الشكلي

تبحث هذه المحركات في أشكال معينة من وسائط المعلومات مثل الصوت أو المادة المصورة أو لقطات الفيديو. وهي تعتمد في بحثها لهذه الملفات على رموز لغة النص الفائق html ورموز meta tag، وتستطيع بذلك التعرف إلى URL وعمل رابط له داخل قواعد بياناتها. وهي تنقسم إلى نوعين من المحركات هي:

---

(١) http://www.ajkids.com

(٢) http://www.Yahooligans.com

(٣) http://news.Google.com/

(٤) http://news.yahoo.com/

(٥) http://news.Altavista.com/

(٦) http://wwwfroogle.com

(٧) http://www.msnshopping.com

(٨) http://www.yahooshopping.com

١- **محركات بحث عامة** : ويطلق على مثل هذا النوع من محركات البحث مصطلح General Media Search Engines التي تجمع مختلف أنواع الوسائط التي تحمل المعلومات.

ومن نماذج محركات بحث الوسائط المتعددة :-

محرك البحث Alltheweb [١]

محرك البحث Altavista [٢]

٢- **محركات بحث متخصصة** : وهي محركات متخصصة في شكل معين من الوسائط مثل الفيديو حيث تقوم باقتناء وتكشيف وبحث ملفات الفيديو من خلال سمات ملف الفيديو، الإلكترونية Content Based Video Retrieval. ثم تعمل على الترتيب الإلكتروني لملف الفيديو كمقاطع مع ترتيب مشاهد الفيلم التي تدور حول مكان أو زمان واحد، كما يتم ترتيب لقطات كل مشهد أيضا ثم وصف كل وحدة من هذه الوحدات بحسب بداية الفيديو إلى نهايته. هذا بجانب الوصف الببليوجرافي لملف الفيديو. وكذلك محركات بحث المواد المسموعة، حيث تعتمد على الروابط التي تضعها مع أسماء الأغاني أو ما شابهها من الملفات الأخرى. كما يغلب على هذا النوع العمل بأسلوب التصفح من قوائم تبدأ عادة بالمغني ثم الأعمال التي أنتجها، ومن ثم يعتمد الاسترجاع على ربط الأغاني بأسمائها. وكذلك محركات بحث المواد المصورة وهي كثيرة وسيتم تناولها بالتفصيل في هذا الفصل.

ومن نماذج محركات البحث المتخصصة

محرك بحث صوت فقط The Music Finder [٣]

محرك بحث صور فقط Picsearch [٤]

محرك بحث صور وصوت فقط Lycos Pictures & Sounds [٥]

---

(١) http://multimedia.Alltheweb.com

(٢) http://www.Altavista.com/mutimedia/default?qbmode=

(٣) http://www.music-finder.net/

(٤) www.Picsearch.om

(٥) http://multimedia.lycos.com/

محرك بحث صوت وفيديو فقط Singingfish[١]

٤/١ محركات بحث المادة المصورة

تعد محركات المادة المصورة أو المادة المصورة image على اختلاف أشكالها مثل الرسومات والمادة المصورة الفوتوغرافية والكاريكاتير، والخرائط، وغيرها .. واحدة من محركات البحث في الوسائط المتعددة. ولا يختلف بحث المحركات في ملفات الوسائط بصفة عامة وملفات المادة المصورة بصفة خاصة عنها في ملفات النصوص، ففي كلتا الحالتين تعتمد محركات البحث على بنية الوثيقة المكتوبة بلغة HTML في التعرف إليها وضمها إلى قاعدة بياناتها، ويتعدى الأمر ذلك إلى البحث عن المصطلحات الكشفية والكلمات المعبرة عن الموضوع بحسب ما توجد عليه صفحات المواقع. فيمكن لمحركات البحث استخدام رموز (meta tags) المعيارية والتي يدون فيها مصمم الصفحة كل المصطلحات المعبرة عن موضوع الموقع. وفي حالة بحث موقع الوسائط المتعددة. فإن محركات البحث تعتمد على المعلومات الواصفة لموضوع وشكل المادة المرئية أو المسموعة داخل الملف الميتاداتا.

١/٤/١ مفهوم محرك بحث المادة المصورة

لقد طرحت Paula Berinstein بعض الملامح الأساسية الخاصة بمحركات بحث المادة المصورة والتي يجب توافرها في المحرك الكفء وهي:

١. أن يسمح بالبحث عن المادة المصورة باستخدام الكلمات المفتاحية واستخدام عناصر المادة المصورة مثل التاريخ والمحتوى والمصور .Photographer
٢. أن يسمح بالبحث باستخدام عناصر المادة المصورة من لون وشكل والعناصر الأخرى.
٣. أن يبحث قواعد البيانات الداخلية للمواقع الأخرى.
٤. أن تعرض المادة المصورة كجزء من النتائج المسترجعة.
٥. أن يسمح بعرض حقوق النشر وحقوق الاستخدام وملكية المادة المصورة المادية.

---

(١) http://search.singingfish.com/sfw/home.jsp

وتقدم بعض محركات بحث المادة المصورة طرقًا خلاف البحث بالكلمات المفتاحية. وأهم هذه الطرق هي تقسيم المجموعات إلى قوائم مصطلحات تتدرج من الأعم إلى الأكثر تخصصًا، ويرتبط بكل مصطلح أو موضوع مجموعة من المادة المصورة التي يتم عرضها فور النقر على رأس الموضوع وأفضل ما يمثل ذلك هو موقع Corbis[١] .

وتعمل محركات بحث المادة المصورة كغيرها من المحركات على إضافة مواقع هذه المادة إلى قاعدتها وتكشيفها ثم طرحها للبحث أمام المستفيدين؛ إلا أن الفارق الملاحظ في أدوات بحث المادة المصورة هو أنها ليست في مجملها محركات بحث بالمعنى السابق، وإنما قد تكون في أحيان كثيرة قواعد بيانات للصور لا تعمل على تزويد مقتنياتها من الويب وإنما تتبع شركات لها مجموعات صور خاصة وبذلك يكون التزويد كما هو عليه الحال في مقتنيات المكتبات، وتحمل قواعد بيانات المادة المصورة مئات الملايين من هذه المادة، أي أن الاختلاف هو في غياب برنامج spider or crawler. والفارق بين الاثنين كبير حيث يقتصر دور قاعدة البيانات على إستراتيجيات البحث فقط في إدخال المادة المصورة كتسجيلة تكون المادة فيها جزءًا من هذه التسجيلة ولا تقوم بعمليات الإبحار على الويب وجمع ملفات المادة المصورة.

ولقد لاحظ الباحث أن تجميع وتكوين أدوات بحث المادة المصورة يكون على نوعين أساسين هما:-

١. في حالة ما يكون بحث المادة المصورة جزءًا من عمل أداة بحث المادة المصورة، فإن أداة البحث في هذه الحالة تتعامل مع ملفات المادة المصورة كغيرها من ملفات النصوص بأن تستخدم برنامج الزاحف أو العنكبوت في تجميع وإضافة ملفات المادة المصورة إلى قواعد بياناتها.

٢. في حالة ما تكون أداة البحث متخصصة في بحث المادة المصورة فقط، فإن الإضافة هنا لا تكون باستخدام برامج الزاحف أو العنكبوت فقط، وإنما باستخدام العنصر الإنساني بالإضافة إلى قواعد البيانات. كل ذلك يأتي من

---

(١) www.Corbis.com <http://www.Corbis.com

أن أدوات بحث المادة المصورة أقرب إلى كونها قواعد بيانات للصور أكثر من كونها أداة بحث لصور الويب.

وحيث إن عملية البحث عن المادة المصورة ليست قاصرة فقط على استخدام محركات البحث في الويب، وإنما أيضا قواعد البيانات المختلفة، فإن مصطلح نظم استرجاع المادة المصورة سيكون هو المستخدم في هذا الفصل.

٢/٤/١ نظم إدارة واسترجاع المادة المصورة

تعود تسمية نظم إدارة واسترجاع المادة المصورة بصفة عامة إلى الكيفية التي يتم العمل بها في تنظيم مجموعات هذه المادة ؛ حيث يمكن أن توجد نظم التصفح، وهي تنقسم بدورها إلى التصفح من خلال الكلمات المفتاحية أو المكانز، فضلا عن التصفح من خلال لقطات المادة المصورة ذاتها أو نظم البحث الحر Open Search System؛ إلا أنه يمكن أن نقسمها إلى فئتين هما نظم إدارة المادة المصورة، ونظم استرجاع المادة المصورة:

١/٢/٤/١ نظم إدارة المادة المصورة

لقد أورد معهد الخدمة الفنية الاستشارية للصور Technical Advisory Service for Images المعروف اختصارًا باسم TASI أربعة أنواع عامة لإدارة ملفات المادة المصورة الرقمية وهي كما يلي :-

١/ ١/١/٢/٤ نظم تصفح الملفات

تعمل تلك البرامج على تصفح ملفات أو حافظات المادة المصورة داخل نظام الاسترجاع دون الرجوع إلى تلك المادة ومعالجتها فنيًا قبل العمل على استدعائها؛ حيث لا يستخدم البرنامج سوى بيانات الفهرسة الأساسية التي تتعلق بوجود ملف المادة المصورة الرقمي على جهاز الحاسب الآلي. ومثال لذلك برنامج ACDSee على اختلاف إصداراته، ويعمل هذا البرنامج في بيئة الويندوز Windows. ويقوم البرنامج فقط بعرض مجموعات المادة المصورة داخل جهاز الحاسب على اختلاف أماكن الحفظ على القرص الصلب دون معالجة ذاتية أو إنتاج بيانات عن ملفات المادة المصورة.

٢/١/٢/٤/١ نظم البحث البسيطة

تعمل هذه النظم على توفير الحد الأدنى من المعالجة الفنية لملفات المادة المصورة داخل قاعدة بيانات اختزان ملفات المادة المصورة. حيث تقوم هذه البرامج بالوصول إلى مجموعات المادة المصورة في أماكن حفظها ثم إنتاج وحدات على شكل مصغر منها وإنتاج بعض حقول البيانات التي يمكن من خلالها للمستفيد البحث بها واستدعاء تلك الوحدات المصغرة "اللقطات" ثم استدعاء الوحدات الأصلية للصور إذا ما أراد المستفيد الحصول على أصل ملف المادة المصورة الرقمية.

وتمتاز نظم البحث البسيطة عن نظم التصفح السابقة في إمكانية إنتاج بعض حقول الميتاداتا في عدم وجود معايير ثابتة لإنتاج كافة الحقول الممكنة عن التعريف بملفات المادة المصورة، كما أن هذه النظم لا تعتمد حتى الأن على أي نوع قياسي من أنواع قواعد البيانات التي تناسب العمل مع ملفات المادة المصورة. وتعمل أيضا على تقييم أو وصف وحدة واحدة من داخل مجموعات المادة المصورة داخل البرنامج. وتقوم النظم البسيطة الآن بقراءة بيانات ملفات المادة المصورة القائمة على معايير مختلفة مثل IPTC, DIG35, EXIF and Photoshop document، إلا أن تلك البرامج لم تصل حتى الآن إلى قراءة الحقول الخاصة بصفحات الويب.

٣/١/٢/٤/١ نظم التصفح الكاملة

تعمل هذه النظم بشكل أكثر تقدمًا مع ملفات المادة المصورة ؛ حيث تهتم هذه النظم بالتعامل مع حقول الميتاداتا أكثر من اهتمامها بالمادة المصورة ذاتها وذلك على غير سابقتها من النظم الأخرى. ولذلك فإن هذا النوع من النظم يعمل على إنتاج واجهة بحث ذات خيارات متقدمة ومتعددة لبحث ملفات المادة المصورة. وتقوم هذه النظم باختزان مجموعات المادة المصورة الرقمية في قواعد بيانات قادرة على المعالجة الفنية الدقيقة لملفات المادة المصورة تنظيمًا واختزانًا وبحثًا.

وتعمل هذه النظم على استيراد وحدات المادة المصورة الرقمية من خارج النظام سواء كان ذلك بشكل ذاتي أو من خلال التعامل اليدوي مع ملفات المادة المصورة، ثم تقوم بإنتاج البديل للوحدات الأصلية وهي عبارة عن لقطات يتم

اختزانها بمكان ثابت داخل قاعدة البيانات مع إنتاج رابط لهذه الوحدة يمكن من استدعائها مرة أخرى. والجدير بالذكر أن ملفات المادة المصورة داخل قواعد البيانات هنا تأخذ أسماء مطولة مدعومة البحث من النظام إلى جانب إدخالها داخل حوافظ folders معدة من نظام الإدارة ذات أسماء محررة ومحددة من جانب قاعدة البيانات. وتخدم النظم الكاملة الأعداد الكبيرة من ملفات المادة المصورة خلاف ما يكون عليه الحال داخل النظم السابقة، مما يعني معه ارتفاع تكاليف بناء تلك النظم لما تتسم به من ملامح.

١/٤/٢/١/٤ نظم إدارة المادة المصورة المخصصة

تحتاج هذه النظم إلى تحديد احتياجات المستفيدين من نظام إدارة المادة المصورة الرقمية، وذلك لأنه يتم إعداد البرامج المختلفة التي يحتاجها التعامل مع المادة المصورة الرقمية. وتتنوع تلك البرامج بين برامج الإضافة أو التزويد وبرامج بناء قواعد البيانات وبرامج البحث داخل ملفات المادة المصورة. ويمكن القول جملة أن هذه النظم تنشأ اعتمادًا على احتياجات محددة تستدعي الحاجة معها بناء نظام لإدارة ملفات المادة المصورة الرقمية ذا طبيعة خاصة.

١/٤/٢/٢ نظم استرجاع المادة المصورة

تختلف نظم استرجاع المادة المصورة باختلاف النظم لما هي عليه الأسئلة البحثية من اختلافات، وقد ذكرت Sharon Moore أنواع الأسئلة البحثية التي يتم على إثرها استرجاع ملفات المادة المصورة داخل قواعد البيانات كما يلي(١):-

- استخدام أسلوب التصفح داخل قاعدة بيانات المادة المصورة، وذلك من خلال التصفح المرئي باستخدام اللقطات أو التصفح بالكلمات المفتاحية.
- استخدام كل أو جزء من المادة المصورة موضع البحث لاسترجاع مثيلاتها من داخل قاعدة بيانات المادة المصورة.
- استخدام صفة أو عدة ملامح تجمع بين مجموعة من المادة المصورة المراد استرجاعها؛ كأن يتم استدعاء كل المادة المصورة التي تخص تاريخًا محددًا أو فنانًا محددًا.

(١) Moore, Sharon. Image Retrieval, [w. p.], 2001, Visited at 10/10/2003, at http://tlc.dcs.gla.ac.uk/students/level3/imagedb/node13.html

- استخدام التفاصيل التابعة لملف المادة المصورة " حقول وفقرات الوصف " لعمل المضاهاة مع كلمات البحث المدخلة إلى قاعدة البيانات.

ولقد تعددت الأدوات فيما بينها من حيث طريقة الحفظ وطرق البحث المستخدمة مع المادة المصورة الرقمية، ويمكن تناول أنواع نظم استرجاع المادة المصورة الرقمية حسب طرق البحث التي أشارت إليها Sharon Moore كما يلي[١] :-

١/٢/٢/٤/١ استرجاع المادة المصورة بأسئلة البحث المباشر

يعرف هذا النوع بأنه يقوم على تنظيم واسترجاع المادة المصورة الرقمية من خلال مجموعة الصفات المادية المشتقة من مكونات وشكل المادة المصورة الرقمية، بما يمكن معه بعد ذلك طلب تلك المادة المصورة بطرح إحدى صفاتها المادية في عملية البحث. ويستخدم هذا النوع خصيصًا في قواعد البيانات التقليدية المعتمدة على تسجيلات وحقول بيانات لحفظ المادة المصورة الرقمية. وتنقسم نظم استرجاع المادة المصورة بأسئلة البحث المباشرة إلى نوعين هما :-

- نظم وصف المادة المصورة الرقمية: يعتمد هذا النظام على استخلاص مجموعة الصفات التي تعبر عن مختلف جوانب المادة المصورة الرقمية وتحديد عناصر المحتوى الموضوعي داخلها مع صياغتها في كلمات واصفة حيث يمكن طلب كل صورة داخل النظام بمجموعة الواصفات المعبرة عنها. وعلى ذلك فإن هذا النوع يحتاج إلى جهود كبيرة في عمليات تكشيف المادة المصورة؛ حيث سيحتاج تكشيف المادة المصورة إلى الجهد الإنساني في تكوين المحتوى الموضوعي للصورة... بينما على الجانب الآخر فإن التكشيف الآلي للمادة المصورة يحتاج إلى تقنيات عالية في معالجة اللغة الطبيعية لكي يتم الخروج بالكلمات المفتاحية المعبرة عن محتوى المادة المصورة والشكل المادي من مجموعة الجمل المكونة للسياق المصاحب للصورة في صفحة الويب. ويمكن القول هنا إن الصعوبات التي يواجهها هذا النوع من النظم هي الحاجة إلى الدقة العالية في التكشيف والقدرة على التعبير عن كل موضوع من المحتوى الموضوعي العام للصورة.

- نظم تحليل ملامح المادة المصورة : يبنى هذا النوع من نظم الطلب المباشر

---

(١) ibid.

على استرجاع المادة المصورة الرقمية بالمكونات الأصلية لوعاء المادة المصورة مثل الشكل shape واللون color والبنية texture. حيث تأخذ هذه الملامح قيمًا عددية يمكن التعامل بها مباشرة مع المادة المصورة دون احتياج النظام إلى كلمات واصفة أو نص مصاحب؛ حيث لا قيمة للنص في الاستدعاء. ويعد ذلك ما يسمى في أحيان كثيرة بـ استرجاع المادة المصورة المبني على المحتوى Content Based Image Retrieval. ويمكن أن تكون تلك الملامح سهلة القياس مثل عدد نقاط كثافة اللون الأحمر داخل المادة المصورة أو الشكل الهندسي المكون للصورة. أو أن تكون تلك المقاييس معقدة مثل القواعد الخاصة بتنظيم البيانات الهندسية مثل Geometrical Materials، وعلى ذلك فإن مثل هذه القواعد تعتمد على الكميات والثوابت في استرجاع المادة المصورة الرقمية وهو ما لا يناسب الطبيعة التقليدية بين مستخدمي الويب في الحصول على المعلومات بالكلمات والعبارات البحثية.

١/٤/٢/٢/٢ الاسترجاع بالنماذج المصورة

يعود هذا النوع إلى مفهوم ظهر عام ١٩٨٠ لتنظيم واستدعاء المادة المصورة الرقمية على استرجاع مجموعات المادة المصورة باستخدام مثال بحثي لصورة أو أكثر لاستدعاء المادة المصورة الأخرى داخل قاعدة النظام المطابقة لذلك المثال. ويحاول هذا النظام استخدام مبدأ الاستدعاء بالنص الكامل من خلال اختيار المستفيد لمثال من المادة المصورة يعبر به عن حاجته الموضوعية من المادة المصورة ثم البحث بها داخل المجموعة الموجودة ومن ثم يمكن للنظام حساب قيم اللون والشكل وقيم بنية المادة المصورة موضوع البحث ثم مطابقتها مع مختلف المادة المصورة، ويستخدم النظام تلك المعايير كعامل لقبول أو رفض المادة المصورة المستدعاة من قاعدة البيانات؛ فإذا ما احتوت المادة المصورة المطلوبة على قيم اللون الأحمر كان ذلك معيارًا لقبول أو رفض مجموعات المادة المصورة داخل القاعدة.

وينقسم هذا النظام إلى ثلاثة أساليب للعمل هي:-

- استدعاء المادة المصورة بالنماذج المصورة الخارجية: يقصد بذلك استخدام المستفيد لإحدى المادة المصورة الرقمية من خارج قاعدة بيانات النظام للبحث بها داخل النظام. ويحصل المستفيد على تلك النماذج الخارجية من

مصادر مختلفة مثل تحويل المادة المصورة الفوتوغرافية التقليدية إلى الشكل الرقمي بواسطة أجهزة المسح الضوئي Scanners، أو أن يتم تحميلها من إحدى صفحات الويب ثم إدخالها للبحث بها داخل نظام الاسترجاع. ومن أهم مميزات هذا النظام هو سهولة وصف المستفيدين لاحتياجاتهم الموضوعية من المادة المصورة، من خلال تقديم نموذج مصور لنوع وشكل المعلومات المصورة المطلوبة.

- استدعاء المادة المصورة بالنماذج المصورة الداخلية : يحتاج المستفيد في هذا النموذج إلى استخدام محتويات قاعدة بيانات نظام الاسترجاع لتحديد أفضل نموذج أو صورة رقمية تعبر عن الحاجة الموضوعية، ثم استخدامها في بحث مجموعات المادة المصورة الرقمية الداخلية للنظام. وتأتي ميزة هذا النوع في أنه لا يرهق المستفيد بالحصول على نموذج صورة خارجية ثم البحث به داخل النظام، على حين أن هناك وقتًا إضافيًا يستنفده الباحث في إجراء عمليات بحث هذا النظام ذلك لما يحتاجه من وقت في إخراج صورة البحث، ثم يأتي وقت البحث الأصلي.
- استدعاء المادة المصورة بالرسوم التصويرية : يعطي نظام استرجاع المادة المصورة في هذه الحالة إمكانية الرسم التصويري لمحتوى المادة المصورة المراد استدعاؤها من بين مجموعات قاعدة البيانات. ويبنى ذلك في بادئ الأمر على إمكانية تصوير المستفيد لجوانب المحتوى الموضوعي للصور، هذا إلى جانب اختيار الشكل المحدد للصور مع اللون والخصائص الأخرى؛ إلا أن أهم الصعوبات التي يواجهها هذا النظام هي مدى توافر قدرات التصوير والرسم عند المستفيد في الوقت الذي يفتقر إلى تلك القدرة الكثير من المستفيدين.

٣/٢/٢/٤/١ استدعاء المادة المصورة الرقمية بالتصفح

يفترض نظام التصفح هنا احتياج المستفيد إلى صورة محددة أو أكثر قليلا؛ لذا فإن التصفح يعمل على إتاحة الوصول إلى محتويات قاعدة بيانات النظام للتعرف إلى جوانبها الموضوعية المختلفة، وتعطي نظم الاسترجاع لذلك إمكانية التصفح المرئي بالتجول خلال مجموعات صور قاعدة البيانات، هذا فضلا عن التصفح بالمصطلحات المفتاحية.

ويعد التصفح الشريك الآخر للبحث بالكلمات المفتاحية في كل من نظم استرجاع المعلومات على الويب المتمثلة في محركات البحث وأدلة البحث. ويتسم أسلوب التصفح في استدعاء مجموعات المادة المصورة بثلاثة ملامح أساسية هي :-

١) أنه الأسلوب الأكثر تفاعلية بين المستفيد ونظام الاسترجاع؛ حيث يرجع الجهد الأكبر في استدعاء بيانات محددة إلى اختيار وانتقاء المستفيد بما يعطيه القدرة على رؤية جوانب قاعدة البيانات المختلفة.

٢) أن المستفيد هنا هو المحدد لاتجاه البحث وبناء البيانات المسترجعة؛ ذلك لما يتسم به أسلوب التصفح من كلمات تساعد على تغيير اتجاه البحث مثل كلمات السابق، إلى الأمام، إلى المستوى الأعلى، إلى المستوى الأقل.

٣) أن أسلوب التصفح لا يتطلب من المستفيد معرفة مختلف أوامر التعامل مع قاعدة البيانات فضلا عن البعد عن الجوانب المتقدمة في بناء إستراتيجيات البحث لكل نظام على حدة، بالشكل الذي يرهق المستفيد في معرفة أشكال مختلفة لأوامر واحدة أو أوامر مختلفة لنظام واحد.

ويوجد شكلان مختلفان من تصفح قواعد بيانات المادة المصورة هما :-

- **التصفح المرئي** : ويبنى هذا الشكل على تصفح مجموعات صور النظام واحدة تلو الأخرى، مما يعطي أبسط أشكال استرجاع المادة المصورة. وتعد أهم ميزات هذا الشكل هو استخدام المستفيد للعين المجردة في التأكد من وجود صورة بذاتها ضمن مجموعات المادة المصورة أو اختفائها عن النظام. أما ما يشوب هذا الشكل من عيوب فهو حاجة المستفيد إلى تصفح كل مجموعات النظام لإيجاد صورة واحدة بعينها، ويزيد الأمر صعوبة كلما زادت محتويات قاعدة البيانات من مجموعات المادة المصورة الرقمية.

- **التصفح بالمصطلحات المفتاحية**: ويبنى هذا الشكل على تقسيم مجموعات المادة المصورة الرقمية داخل النظام إلى مجموعات كبرى متكاملة في الموضوع ثم تقسيم مجموعة المادة المصورة للموضوع العام إلى مجموعات لموضوعات متخصصة فرعية للموضوع العام. وعلى ذلك فإن شكل التصفح هنا يأخذ شكل القوائم المنبثقة عن قوائم أخرى أعم، وصولا إلى القائمة الرئيسة الأولى. هذا بجانب استخدام تقنية الربط الفائق بين كل مصطلح

دال في كل قائمة ومجموعة المادة المصورة داخل قاعدة البيانات المعبرة عن دلالة هذا المصطلح.

وتعد أهم ميزات استخدام هذا الشكل هو تحرك المستفيد داخل مجموعات من المادة المصورة المحددة حول موضوع بعينه، مما يفيد سرعة الوصول إلى المجموعة المحددة من المادة المصورة.

٤/٢/٢/٤/١ نظم استرجاع المادة المصورة الرقمية المتكاملة

نظرًا لما اتسمت به الأنواع السابقة من نظم استرجاع المادة المصورة الرقمية بالكثير من الملامح التي تختلف فيما بينها بين الإيجابي والسلبي في دقة وكفاية استرجاع المادة المصورة الرقمية. فقد حاولت النظم المتكاملة لاسترجاع المادة المصورة الجمع بين ميزات النظم السابقة مع تلافي عيوبها. وقد تجمع النظم المتكاملة بين شكلين أو أكثر من أشكال الاسترجاع السابقة؛ حيث تعطي إمكانية الاستفادة من خصائص البحث المختلفة، وذلك مثل الجمع بين طريقتي التصفح المرئي والبحث المباشر بالكلمات المفتاحية.

ويمكن القول هنا إنه يوجد تكامل بين مختلف طرق بحث مجموعات المادة المصورة وبعضها البعض وذلك بأساليب مختلفة؛ فالأسلوب الأول لذلك التكامل يأتي في اعتماد نظم استرجاع المادة المصورة على البحث بالكلمات المفتاحية ثم عرض مجموعات المادة المصورة المطابقة للبحث في شكل شاشات نتائج يمكن للمستفيد تصفحها والاختيار فيما بينها. وهذا في الوقت الذي يستخدم فيه النظام أسلوبًا واحدًا فقط هو الكلمات المفتاحية. أما الأسلوب الثاني فهو تنشيط مختلف طرق البحث وإعطاء واجهات بحث مختلفة تتشكل جوانبها كما يوافق طريقة البحث.

٥/١ الخلاصة

تختلف ماهية الويب عن الإنترنت ؛ حيث تعتبر الويب جزءًا من بناء الإنترنت له خصائصه المختلفة عن غيرها من الأجزاء. وتتنوع أساليب البحث على الويب ما بين استخدام أدوات البحث إلى البحث في قواعد بيانات الويب. وقد عمل هذا الفصل على عرض أنواع محركات البحث مع التركيز على محركات بحث الصور ثم الحديث عن نظم إدارة الصور الرقمية.

الفصل الثاني

المادة المصورة في البيئة الرقمية تطبيقاتها وتقنياتها



المادة المصورة في البيئة الرقمية
تطبيقاتها وتقنياتها

٢/٠ التمهيد

إن الحديث عن أنواع المادة المصورة في البيئة الإلكترونية أو الرقمية لا يختلف عنه في البيئة التقليدية، وعلى ذلك فسيكون مضمون هذا الفصل مفهوم المادة المصورة وتطبيقاتها في مختلف المجالات .

٢/١ المادة المصورة من الكهوف إلى الويب

لقد انتهى تسجيل المعلومات في العصر الحديث إلى ما بدأ عليه في العصور السابقة؛ حيث بدأ التسجيل للمعلومات كرموز على الجدران وانتهى الآن إلى التسجيل كرموز أيضا هي لغة الحاسب (النبضات الضوئية) أو النظام الثنائي 10010011. حيث يكون الشكل المحدد للرقم ذا دلالة على معلومات محددة، بل كانت العناصر المصورة أيضا هي مكونات الرسالة التي أرادت الحضارات السابقة توصيلها. إن تسجيل الصور لإفادة معلومات محددة مثل المعارك الحربية لم يكن يتم من خلال آلات التصوير كما هو الآن وإنما كان السبيل الوحيد هو الرسم على الحوائط. ويذكر المعهد العلمي للبيانات المصورة Institute for image data research أن المادة المصورة ترجع في تاريخها إلى ما قبل الحضارة الرومانية .

واتخذت المواد المصورة بعد ذلك أشكالا كثيرة في طريقها لإيصال المعلومات معتمدة في هذا على تطور تكنولوجيا البث المرئي للمعلومات. فأصبح هناك بجانب آلات التصوير التقليدية، آلات التصوير بأشعة إكس وتحت الحمراء وفوق البنفسجية للأراضي الطبيعية، ثم بعد ذلك آلات التصوير الميكروسكوبي لأغراض البحث العلمي، وأخيرًا الأقمار الصناعية. وتعمل الأقمار الصناعية على

التقاط صور لشكل الأرض الخارجي أو التقاط صور على عمق أبعد من القشرة الأرضية، ذلك خلاف الصور التي تتعلق بالفلك والنجوم .

أما في مجال التخزين فقد دخلت الصور إلى الحاسبات عام ١٩٦٥م على يد العالم Ivan Sutherland الذي بدأ مشروعه في رقمنة الوسائط المصورة، وأخذت هذه الوسائط طريقها ببطء مع الحاسب إلى منتصف الثمانينات الذي شهد ألعاب الكمبيوتر، مما كان له أكبر الأثر في تحسيب هذا النوع من الوسائط. ثم استخدمت تقنيات الأقراص الضوئية في تخزين الصور الطبيعية والفنية أو التي تختص بالحيوانات والأشكال الطبيعية في المكتبات، ولقد لاحظ الباحث أن الاسترجاع داخل هذه القواعد غير المحدثة يتم من خلال التقسيم الموضوعي أو التصفح أيضًا من خلال البحث بالكلمات المفتاحية المضافة إلى ملف الصورة. هذا إلى جانب انتشار وتداول قواعد بيانات الصور على الأقراص الضوئية بين الأفراد حيث يمكن نقل كمية كبيرة من مجموعات الصور بطريقة بسيطة إلى جانب إمكانية التحكم في الاستخدامات المختلفة للصور على الأقراص. ومن أهم قواعد البيانات المتاحة على أقراص ضوئية هي؛ photosphere وهي تحمل مجموعة من الصور الفردية والفنية، وقاعدة the stock solution وهي تحمل مجموعة من الصور والرسومات تصل إلى ٢٠٠,٠٠٠ صورة. ويمكن القول إن أنواع المادة المصورة الثابتة جاءت مختلفة تبعًا للفترات الزمنية المختلفة التي سادتها أشكال مختلفة مثل الرسومات واللوحات ثم الصور الفوتوغرافية التقليدية ثم الصور الفوتوغرافية الرقمية.

ثم شهدت هذه الوسائط بعد ذلك الطفرة الحقيقية في الدخول إلى المجال الرقمي عن طريق الويب في عام ١٩٩٣م إلى أن قدر عددها في عام ١٩٩٧م ما بين ١٠ إلى ٣٠ مليون وسيط، ولقد تضاعف هذا العدد عشرات المرات في بضع سنوات حيث يقدر عددها على موقع مثل Corbis وهو موقع واحد فقط من مواقع المادة المصورة على الويب بحوالي ١٢٠ مليون مادة، يمكن البحث في ٦٠ مليون منها على الخط المباشر.(١) وتشير الدراسات الخاصة ببحث المادة المصورة

---

(١) Eakins, John P and Graham, Margaret E. Content-based Image Retrieval A report to the JISC Technology Applications Program, the institute, 1999, visited at 24/9/2002, http://portal.acm.org/citation.cfm?id=614839

الرقمية أن أعداد ملفات الصور على الويب تزيد بمعدل مليون أو أكثر من الملفات، وأن حجم المادة المصورة الرقمية على الويب يزيد على ٣ تيرابايت[١]، مما يعكس معه الحجم المتزايد من ملفات المادة المصورة والذي يحتاج معه أيضًا استعدادًا أكبر لمحركات البحث الداعمة لبحث هذا الوسيط الرقمية[٢] .

وتختلف أنماط الحديث عن المادة المصورة بحسب المراحل التاريخية التي عاشتها الصور في أداء مهمة الاتصال والتواصل بين الأفراد. ففي مراحل عدة منها؛ كان الحديث عن الرموز المصورة التي وجدت على جدران المعابد والآثار القديمة، وهي البداية الفعلية في استخدام المادة المصورة لنقل المعلومات. أما المرحلة الثانية فتمثلت في اختراع آلات التصوير الفوتوغرافي، وبدأت على ذلك مرحلة خاصة تبتعد بها عن الأنواع الأخرى من المواد المصورة مثل اللوحات والرسوم المختلفة... وغيرها. ثم تطورت آلات إنتاج الصور الفوتوغرافية إلى نقطة التلاقي مع استخدامات الحاسب ليبدأ Ivan Sutherland عام ١٩٦٥م مشروعًا هدف من خلاله إلى رقمنة المادة المصورة لاستخدامها في تطبيقات الحاسب. ثم تطورت مشروعات المادة المصورة الرقمية إلى أن أضحت الجانب المميز لخدمة الويب عن غيرها من خدمات الإنترنت. ومنذ عام ١٩٩٣م بدأ هذا الوسيط من خلاله وجوده على الويب رحلة البحث عن التقعيد لعمليات المعالجة الفنية والاختزان والبحث داخل مواقع الويب، إلى أن أصبحت مشكلة التعامل بين المادة المصورة ومواقع البحث (محركات البحث) مشكلة تحتاج إلى حل فعال لتحقيق الكفاية المرجوة في نظم استرجاع المادة المصورة على الويب.

ويأخذ الحديث عن تناول الإنتاج الفكري المادة المصورة الرقمية أكثر من شكل بحسب الجوانب التي تتناولها، كما سبق الإشارة لذلك. ويمكن ذكر بعض الجوانب مثل :-

- المادة المصورة الرقمية من منظور علم الحاسب في الاختزان والتكوين.

---

(١) التيرابايت هي وحدة من وحدات قياس الذاكرة داخل الحاسبات الكبيرة MAIN FRAME و هي تبلغ ١٠٠٠ جيجا بايت تقريبا.

(٢) Goodrum, Abby A.. Image information retrieval: an overview of current research, informing science, 2000, visited at 1/10/2003, http://64.233.161.104/search?q=cache:BRhAug3cfd

- المادة المصورة الرقمية في قياس آلات التصوير الرقمية وتقنية الكثافة.
- المادة المصورة الرقمية كنوع من أنواع التصميم الفني.
- المادة المصورة الرقمية التجارية (الدعاية والتسويق).
- المادة المصورة الرقمية البحثية والتعليمية.
- المادة المصورة الرقمية في علم الاتصال.
- المادة المصورة الرقمية كوسائط للمعلومات المرئية (استرجاع المعلومات).

٢/٢ تطبيقات المادة المصورة

١/٢/٢ التطبيقات العامة للمادة المصورة

تحمل أوعية معلومات الصور الكثير من التطبيقات التي يمكن أداؤها على مستويات مختلفة وفي نواحٍ متعددة. وقد ذكرت جمعية الخدمة الفنية الاستشارية للمادة المصورة TASI أن هناك مستويات مختلفة لاستخدام وسيط معلومات الصور ذكرها Ashford and Graham فيما يلي:-[١]

١) عمل الإيضاحات Illustration: حيث تصاحب المادة المصورة النصوص التي تحكي موضوعًا محددًا ويكون للصورة الدور في توضيح المعنى الذي يحمله النص.

٢) معالجة المعلومات: حيث تستخدم المادة المصورة التي تحتوي على موضوعات أو مفاهيم للتعرف إلى سمات هذا الموضوع مثل تشخيص الأمراض أو الرسومات الهندسية.

٣) بث المعلومات Information dissemination: وهو منظور علم المعلومات في أن المادة المصورة ذاتها وسيط معلومات.

٤) التعليم: وهو ما تقدمه المدارس والجامعات ومراكز البحث العلمي من مادة مصورة تستخدم كوسائل تعليمية.

٥) توليد الأفكار Generation of ideas: وهو توظيف المادة المصورة في توليد مجموعة جديدة من الأفكار أو الحقائق.

---

(١) Searching for and Retrieving Digital Images. TASI Technical Advisory Service for Images, 2003, visited at 25/4/2003, http://www.tasi.ac.uk/advice/using/ppt.html

٦) الأغراض الجمالية Aesthetic value: ويكون الاهتمام هنا بالجانب الظاهري المادة المصورة لتوظيفها في أعمال الديكور.

٧) حجية الصورة Emotive/Persuasive: وهي استخدام المادة المصورة كدليل إثبات للتأكيد على حدث أو موقف محدد، هذا فضلا عن جانب الاحتفاظ بمواقف خاصة بالذاكرة الخارجية للإنسان.

٢/٢/٢ التطبيقات العلمية للمادة المصورة

تتعدد احتياجات المستفيدين إلى المادة المصورة وتختلف حسب طبيعة التعامل معها، بداية من الاستخدام الشخصي كذاكرة دائمة خارج الذهن تسجل الأحداث المتتابعة في حياة الإنسان إلى استخدامات الصحف والهيئات العلمية. ومن المهم أولا التعرف إلى تطبيقات هذا النوع من المادة بصفة عامة وتطبيقات الصور الفوتوغرافية بصفة خاصة، ذلك لأن الحاجة إلى تنظيم وسيط معلومات معين يعني كثرة تطبيقات هذا الوسيط واختلاف طبيعة المستفيدين منه. وحري بالذكر هنا أن الباحث يرمي إلى الفترات التاريخية التي جاءت بعد ابتكار آلات التصوير أو تسجيل المعلومات المصورة. على أن الحديث هنا أيضا لا يقتصر على الصور الفوتوغرافية فقط وإنما على كل ما يندرج أسفل مصطلح "Image" مثل الملصقات والشفافيات والرسوم البيانية والخرائط... وغيرها من الوسائط المصورة الثابتة. وكل ذلك من الممكن أن يحمل معلومات ذات قيمة عالية في إعطاء حقائق أو نتائج تؤثر في البحث العلمي. ويفضل الباحث عرض تطبيقات المادة المصورة حسب المجالات العلمية المختلفة، وذلك كما يلي:-

١/٢/٢/٢ في مجال دراسة التاريخ

دائما ما تكون الصور أصدق كثيرًا من الكلمات في وصف الأحداث السابقة، ولأن مجال دراسة التاريخ يعتمد على المادة بشكل رئيس في الحصول على المعلومات اللازمة لاختبار الفرض التاريخي، فإنه يحتاج إلى الوسائط الخالية من الزيف للحصول على أصدق المعلومات. وتعد الصور شاهدًا على صدق أو كذب الكثير من ظواهر الماضي.

ولا تقتصر تطبيقات المادة المصورة في مجال التاريخ فقط على الاستخدام في إثبات الحقائق أو كمصادر لدراسة الظواهر التاريخية. فهناك الكثير من الهيئات المتخصصة في حفظ التراث والتاريخ تعمل الآن على جمع الصور التاريخية في بيئة

الويب ومن ثم حفظها وعرضها ؛وذلك من خلال مواقع المتاحف والجمعيات المهنية في مجال التاريخ. وعلى هذا فإن الكثير من محركات بحث المادة المصورة تعمل على الاستفادة من مثل تلك المجموعات التاريخية. وتشير دراسة David Mattison إلى تعدد مواقع الويب الداعمة للصور التاريخية والتي تتبع منظمات لرعاية المواد المصورة التاريخية [١] مثل جمعية المكتبات الفنية في أمريكا الشمالية The Art Libraries Society of North America التي تم إنشاؤها في عام ١٩٧٢ [٢]، وجمعية الصور الفوتوغرافية التاريخية في كندا the Photographic Historical Society of Canada [٣]، والجمعية الملكية للصور الفوتوغرافية the Royal Photographic Society. [٤]

ومن نماذج قواعد بيانات المادة المصورة في مجال التاريخ على الويب

قاعدة بيانات Open [٥]

قاعدة بيانات Seaborg [٦]

٢/٢/٢/٢ في مجال الهندسة والتصميم المعماري

تعمل المادة المصورة هنا على تقريب وجهات النظر في أنواع التصميمات المعمارية والإنشائية والديكور، وكذلك الرسومات الميكانيكية والكهربائية، ويتم حفظها لأغراض عدة منها التعليمي، والمراجعة، وإثبات وقائع تخص المواقع المعمارية. كما أن للمادة المصورة استخدامًا حيويًا آخر داخليًا يتمثل في أدوات الاختيار (كتالوجات) للأدوات الهندسية أو الآلات في الهندسة الميكانيكية أو أنواع لتصميمات في الهندسة المعمارية والديكور والأثاث... الخ. وتسهم هذه الكتالوجات في تعريف المهندسين بآخر ما توصلت إليه التصميمات الهندسية في مجالهم .

ومن نماذج قواعد بيانات المادة المصورة في مجال الهندسة على الويب

قاعدة بيانات Sussex.ac [٧]

---

(١) Mattison, david. Images of History on the Web. information today,inc, 2002, visited at 15/10/2003,:http://www.infotoday.com/searcher/may02/mattison.htm

(٢) http://www.arlisna.org/

(٣) [http://www.rps.org

(٤) http://www.open.org/~library/fotofind.html

(٥) http://seaborg.nmu.edu/earth/Life.html

(٦) http://www.afp.com/arabic/products/?pid=image

(٧) http://www.lifeart.com

٣/٢/٢/٢ في مجال الإعلام

والحديث هنا عن الإعلام المقروء خاصة والمتمثل في مجموعات الصحف والمسلسلات، التي تحوي الملايين من الصور الفوتوغرافية التي تعبر عن الأحداث الجارية في شتى المجالات. فالأرشيف الصحفي الذي يضم ملايين المقالات والصور ليس فقط للحفظ، بل لاسترجاع هذه الوسائط عند الحاجة إلى معلومات محددة. وتتجه المؤسسات الإعلامية حاليًا إلى تحويل هذه الأرشيفات إلى رقمية يمكن أن يستفيد منها الصحفيون من أي مكان في العالم لتدعيم مقالاتهم.

ومن أشهر نماذج قواعد بيانات المادة المصورة في مجال الإعلام على الويب

قاعدة بيانات وكالة الأخبار الفرنسية AFP[١]

٤/٢/٢/٢ في مجال الطب

يعتمد مجال الطب بصورة أساسية في البحث والتشخيص وتبادل المعلومات على استخدام الصور الفوتوغرافية وصور الأشعة. وتوجد الآن الكثير من قواعد البيانات على الإنترنت تتخذ عناوين متنوعة تختص جميعًا بالصور في مجال الطب لأغراض التعليم، وهذه المواقع تحتوي على رسومات التشريح الخاص بأعضاء الجسد المختلفة. ومن أشهر نماذج قواعد بيانات المادة المصورة في مجال التعليم الطبي على الويب قاعدة بيانات Life Art[٢]

٥/٢/٢/٢ في مجال القانون وعلم الجريمة

لعل أبرز ما يمثل استخدام المادة المصورة الرقمية في هذا المجال هو حفظ واسترجاع صور الأشخاص المطلوبين للعدالة، وصور بصمات اليد والأصابع يرافقها مجموعة من الحقول الواصفة لصاحب تلك البصمات. وتستخدم هيئات الشرطة حاليا الحاسبات في إعداد قواعد بيانات عن الأشخاص الممنوعين من السفر، وهي تضم تسجيلات ولكن بحقول مختلفة حسب طبيعة المعلومات المراد تنظيمها واسترجاعها فيما بعد.

ومن نماذج قواعد بيانات المادة المصورة في مجال القانون على الويب

---

(١) http://fotosearch.com

(٢) http://catalog.lapl.org/a_photo.html

قاعدة بيانات fotosearch[1]

قاعدة بيانات catalog.lapl[2]

٦/٢/٢/٢ في مجال الجغرافيا والسياحة:

يعد مجال الجغرافيا من أهم المجالات التي تعتمد على المادة المصورة بكافة أشكالها، وبصفة خاصة الخرائط والرسومات الطبوغرافية. وتعرف قواعد البيانات الجغرافية عادة باسم Geographical Information System واختصارا GIS. وقد اعتمدت وكالة الفضاء الدولية ناسا NASA على قاعدة بيانات للصور الرقمية المتخصصة في عرض المعلومات عن كوكب الأرض وتحيط هذه القاعدة بكل الصور التي تغطي سطح الأرض بما فيه الجبال والمحيطات والبراكين وأماكن الزلازل أو المعلومات عن الصخور المكونة لكل هذه الأشكال[3]. وتمتد استخدامات الصور الجغرافية ولا سيما المناظر الطبيعية إلى مجال السفر والسياحة التي تمثل للكثير من الدول مصدرا رئيسًا للدخل.

ومن نماذج قواعد بيانات المادة المصورة في مجال الجغرافيا والسياحة على الويب

قاعدة بيانات accuweather.[4]

قاعدة بيانات torro[5]

٧/٢/٢/٢ في مجال التجارة :

يرتبط بهذا المجال الدعاية والإعلان التي تعتمد على الصور والرسومات الخاصة بالسلع. وتحاول الشركات المنتجة على اختلافها تسويق منتجاتها وتستخدم في ذلك الصور الفوتوغرافية المعبرة عن قيمة المنتج وأهميته لجذب المستهلكين إليه.

ومن نماذج قواعد بيانات المادة المصورة في مجال التجارة على الويب

قاعدة بيانات usda[6]

---

(١) http://www.visibleearth.nasa.gov/

(٢) http://accuweather.ap.org/

(٣) http://www.torro.org.uk/photos.htm

(٤) http://www.usda.gov/oc/photo/agri11.htm

(٥) Graham, Margaret E. THE DESCRIPTION AND INDEXING OF IMAGES:

(٦) Report of a survey of ARLIS members, Institute for IMAGE DATA Research1998/99, visited at 6/10/2003, http://www.unn.ac.uk/iidr/ARLIS/.

٨/٢/٢/٢ في مجال التعليم والبحث العلمي

تعد المادة المصورة أداة مهمة جدا وفعالة في مجال التعليم، فتستخدم الصور والرسومات في مرحلة رياض الأطفال ومحو الأمية، وتستخدم الوسائل التعليمية للطلاب في المدارس، وفي مجال البحث العلمي تمثل الصور الملتقطة بالأقمار الصناعية والميكروسكوبية عناصر أساسية مكملة للنص. وقد أكد مسح لمعهد أبحاث البيانات المصورة Institute for Image Data Research على أن الصور تخدم بشكل كبير العمليات التعليمية والبحثية، وذلك من خلال استقصاء آراء عينة من اختصاصيي المكتبات بلغ عددها (١٥٣) اختصاصيًا.[(١)] حيث مثلت المجالات التعليمية نسبة ٦٧٪ من استخدام الصور في المناقشات العلمية والمحاضرات والبحث العلمي والأطروحات الأكاديمية. ومن نماذج قواعد بيانات المادة المصورة في مجال علم الحيوان قاعدة بيانات Animal Pictures Archive.[(٢)]

وخلاصة القول فالمادة المصورة هى الوجه الآخر من العملة المكمل للمادة النصية. ولقد عرضت جمعية الخدمة الفنية الاستشارية للصور القيمة المضافة التي تضيفها الصور للمادة النصية كالآتي:-[(٣)]

- إن وجود صور داعمة للتعبير عن محتوى النص المصاحب يمكن أن تعبر بمفردها عما يهدف إلى عرضه هذا النص.
- إن الصور يجب أن تتطابق مع النصوص المصاحبة لها في الهدف، وإلا فإن النص يمكنه التعبير في حد ذاته عن الموضوع باستخدام الكلمات.
- إن وجود الصور داخل النصوص لا يدعم فقط فهم المحتوى وإنما يساعد المتعلمين أيضا على تذكره.
- إن الصور بطبيعتها المصورة يمكنها أن تدعم قوة التخزين والاسترجاع داخل الذاكرة.

---

(١) http://www.animalpicturesarchive.com/

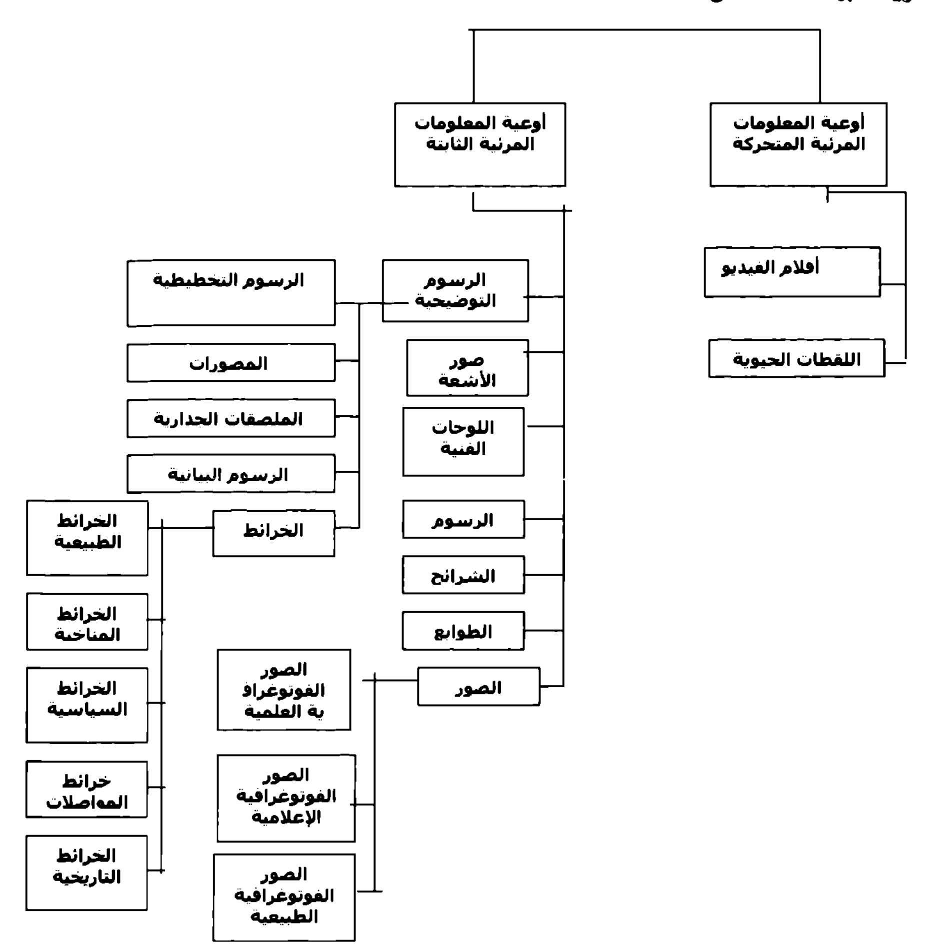
(٢) DIMP: The Use of Images to Support Instruction and Presentation. TASI Technical Advisory Service for Images, 2003, visited at 11/12/2003, http://www.tasi.ac.uk/advice/using/ppt_case.html

(٣) شعبان عبد العزيز خليفة. المواد السمعية و البصرية و المصغرات الفيلمية في المكتبات و مراكز المعلومات / شعبان عبد العزيز خليفة، محمد عوض العايدي- . ط٢، مزيدة و منقحة -. [القاهرة]: مركز الكتاب للنشر، ١٩٩٧. ص٣٣، ٤٣.

- إن الصور في بعض الأحيان يمكنها أن تحل بدلا من النصوص في حالات التعبير عن المعلومات غير الشفوية المراد إيصالها من شخص إلى آخر.
- إن الصور ذات المحتوى الأقرب إلى النص يمكنها أن ترفع من كفاية المتعلمين غير المهرة في التعامل مع محتوى وسيط المعلومات ككل.

٣/٢ أنواع المواد المصورة

يمكن تقسيم المادة المصورة إلى قسمين أساسين هما المادة المصورة المتحركة، والمادة المصورة الثابتة. ويتفرع تحت كل منها أشكال عدة، وفيما يلي تعريف بهذه الأشكال:-



الشكل رقم (١/٢) يوضح مكان الصور الفوتوغرافية الرقمية بين أنواع المواد المرئية

٢/٣/١ المواد المصورة المتحركة

يقصد بالمادة المصورة المتحركة مجموعة الصور التي تتركب من لقطات متتابعة وتعطي إيحاءً بالحركة فور عرض هذه اللقطات بسرعة أكبر من حركة حدقة العين. وأنواع المواد المصورة المتحركة كما يلي:-

٢/٣/١/١ الأفلام

هي "مجموعة من اللقطات مصورة بطريقة متصلة ومرتبة رأسيا وأثناء العرض تعطي الإحساس باتصال الحركة. وقد ينطبق هذا التعبير على الأفلام الصامتة وكذلك على الأفلام الناطقة التي قد يعبر عنها أحيانا بالتعبير الأمريكي "motion pictures"(١) . ويتم تقسيم هذه الأفلام إلى نوعين بحسب نوع الفيلم والمادة المصنوع منها الفيلم. النوع الأول هو الأفلام، وقد كانت هذه الأفلام صامتة ثم طورت إلى الأفلام الناطقة وتوجد أنواع منها تختلف باختلاف الحجم وصناعة الأفلام. أما النوع الثاني من الأفلام المتحركة فهو أشرطة الفيديو، والذي يصنع من مادة أكسيد الكربون. ويحتوي هذا النوع على أشكال مختلفة للاستخدام. وفي البيئة الرقمية تأخذ الأفلام طابعًا خاصًا يجعلها تختلف عن الأفلام التقليدية؛ ذلك من ناحية البحث والتنظيم والتنزيل أو التحميل. كما أن سرعتها تعتمد على الأجهزة المساعدة على العرض السريع، وسرعة التحميل، والمعالجات، ونوع الاتصال بالإنترنت.

ومن نماذج قواعد البيانات التي تتيح أفلامًا على الويب:-

- قاعدة بيانات Rock-n-reel(٢)
- قاعدة بيانات Tdfilm(٣)
- قاعدة بيانات Webmovie(٤)

٢/٣/١/٢ الرسوم الحيوية

تعد الرسوم الحيوية هي نتاج الاعتماد على الحاسبات في التصميم والتي تعرف باسم "الجرافيك Graphic" . والرسوم الحيوية تستخدم مع الموسوعات

---

(١) http://www.rock-n-reel.com/

(٢) http://www.tdfilm.com/

(٣) http://www.webmovie.com/

(٤) http://www.animationusa.com/

الإلكترونية والأغراض التعليمية وغيرها من أعمال تصميمات المباني. ويوجد من هذا النوع من المواد المصورة المتحركة نوعان أساسيان هما الرسوم ثنائية البعد 2D ، والرسوم ثلاثية الأبعاد 3D. وعلى ذلك فإن المواد المصورة المتحركة منها ما يرتبط وجوده بالحاسب مثل الرسوم الحيوية، ومنها ما يمكن للمكتبة التقليدية اقتناؤه وعرضه بواسطة أجهزة خاصة لذلك.

ومن نماذج قواعد البيانات التي تتيح رسومًا حيوية على الويب:-

قاعدة بيانات Animationusa[(١)]

قاعدة بيانات Animationconnection[(٢)]

قاعدة بيانات Allanimation[(٣)]

٢/٣/٢ المواد المصورة الثابتة

تتعدد الأنواع أسفل هذه الفئة، ويطلق على كل هذه الأنواع في الغالب تسمية image على اعتبار أنه اسم الأشكال المصورة الثابتة. وتأخذ محركات بحث الصور مصطلحًا ثابتًا Image Search Engines، وتبحث في الرسوم والصور الفوتوغرافية واللوحات الفنية والخرائط والرسومات البيانية وغيرها. ويكون الاختلاف بين هذه الأشكال في الحقول المميزة لشكل عن آخر مثل مقياس الرسم للخرائط، ونوع الشكل للصورة الفوتوغرافية format. وأنواع المواد المصورة الثابتة كما يلي:-

٢/٣/٢/١ الرسوم التوضيحية

تعد الرسوم التوضيحية أقدم نوع استخدمه الإنسان في إيصال الأفكار، وهي تعني مجموعة المواد البصرية التي تحاول إيصال أفكار أو حقائق محددة من خلال الرسوم والتعليقات اللفظية. وتتعدد أنواع الرسوم التوضيحية إلى ما يلي:-

- **الرسوم التخطيطية** : هي مجموعة من الخطوط المستقيمة والمنحنية التي تهدف إلى توصيل حقائق وتقديرات محددة. وتستخدم الرسوم التخطيطية بجانب الصور الفوتوغرافية في العملية التعليمية؛ إلا أن الصورة الفوتوغرافية

---

(١) http://www.animationconnection.com/

(٢) http://www.allanimation.com/

(٣) http://www.free-graphics.com/

أكثر استقرارًا في ذهن المتلقي، حيث إنها شبيهة بالأصل خلاف ما تكون عليه الرسوم التخطيطية.

ومن نماذج قواعد البيانات التي تتيح رسومًا تخطيطية على الويب:-

قاعدة بيانات Free-graphics (١)

- **المصورات**: وهي تمثيل الشيء الأصلي؛ إلا أنها تهمل العناصر غير الأساسية، وهي خلاف الرسوم التخطيطية في أنها تعتمد على الصورة أكثر من الرمز .

ومن نماذج قواعد البيانات التي تتيح مصورات على الويب:-

قاعدة بيانات Fotosearch (٢)

- **الملصقات الجدارية**: وهي لوحات ذات مكونات مختلفة من الرسومات، والصور الفوتوغرافية تهدف إلى جذب الانتباه إلى شيء محدد أو الإعلام عن شيء معين. وتستخدم هذه الملصقات في الحملات التسويقية وغيرها. والملصقات الرقمية أصبحت ذات أهمية كبرى للشركات القائمة على التسويق على الويب، مما جعل قواعد بيانات الصور تهتم بضم هذا النوع من المواد المصورة.

ومن نماذج قواعد البيانات التي تتيح ملصقات على الويب:-

قاعدة بيانات Free-graphics (٣)

- **الرسوم البيانية**: وهي الرسوم التي تحاول إيصال معلومات عددية أو كمية عن نشاط معين بتمثيل الأرقام والبيانات التي تحويها الجداول في علاقة بين متغيرين. ويعتبر هذا النوع أهم الأنواع المصورة الثابتة في الأغراض البحثية، كما أن قواعد بيانات المادة المصورة الرقمية الآن تضم أعدادًا هائلة من الرسوم البيانية عن مختلف المجالات، مما جعلها ملاذ الباحث في الحصول على ما يدعم به كلامه عن موضوع بحثي معين .

ومن نماذج قواعد البيانات التي تتيح رسومًا بيانية على الويب:-

---

(١) fotosearch.com

(٢) http://www.free-graphics.com/

(٣) http://www.free-graphics.com/

قاعدة بيانات Free-graphics [١]

قاعدة بيانات Ditto [٢]

- **الخرائط**: تمثل الخرائط أداة مهمة في توضيح الحقائق والمعلومات وذلك بالاعتماد على الرموز والإشارات التي تقترب إلى درجة التقعيد. وتوجد أنواع عدة للخرائط منها الخرائط الطبيعية، والسياسية، والمناخية، والسياحية، وخرائط المواصلات، والخرائط التاريخية... الخ. ويستخدم الحاسب الآلي في رسم الخرائط لإضفاء درجة أكبر من الدقة في عمل هذه الخرائط. وهي تستخدم أكثر ما تستخدم في الأغراض التعليمية لدراسة مناهج الجغرافيا والتاريخ. ويمكن تفصيل الخرائط النوعية كما يلي :-

✓ **الخرائط الطبيعية** : ويعرض هذا النوع الأشكال الطبيعية المكونة لسطح الأرض من جبال وهضاب ومنخفضات وبحار ومحيطات. وتستخدم هذه الخرائط ألوانًا ومميزات لكل شكل من الأشكال السابقة يطلق عليها مفتاح الخريطة؛ حيث يمكن عن طريقها تفسير محتويات الخريطة في إعطاء معلومات عن المنطقة التي تحددها الخريطة. ولهذا النوع استخدامات عدة مثل الملاحة وعلوم دراسة سطح الأرض والظواهر الطبيعية المختلفة .

ومن نماذج قواعد البيانات التي تتيح خرائط طبيعية على الويب:-

قاعدة بيانات Naturepl [٣]

قاعدة بيانات Photoartofnature [٤]

قاعدة بيانات Naturepl [٥]

✓ **الخرائط المناخية** : يوضح هذا النوع الظواهر والأحوال المناخية لمنطقة أو مناطق معلومة. واعتمادًا على هذا النوع يمكن دراسة ومتابعة التغيرات المناخية مثل الرياح والضغط والحرارة، ومن ثم يمكن التوقع بتغيرات الطقس

---

(١) http://www.ditto.com/

(٢) http://www.naturepl.com/frmsns.html

(٣) http://www.photoartofnature.com/index.html

(٤) http://www.naturepl.com/frmsns.html

(٥) http://accuweather.ap.org/

والمناخ. وتختلف بدورها العلامات المستخدمة والرموز المكونة لهذه الخرائط اعتمادًا على اختلاف الدلالات المنتظرة من كل نوع من أنواع الخرائط .

ومن نماذج قواعد البيانات التي تتيح خرائط مناخية على الويب:-

قاعدة بيانات Accuweather(١)

قاعدة بيانات Torro.org(٢)

✓ **الخرائط السياسية** : يعنى هذا النوع من الخرائط برسم الحدود بين الدول بما له أهمية في فصل المنازعات السياسية بين الدول وعدم التعدي من دولة على أخرى .

ومن نماذج قواعد البيانات التي تتيح خرائط سياسية على الويب:-

قاعدة بيانات Visoo(٣)

✓ **خرائط المواصلات** : تستخدم هذه الخرائط في عمليات الملاحة الجوية أو البحرية أو البرية، وذلك لتبيان الطرق والمسالك في السفر. ودائمًا ما توجد هذه الخرائط في شركات الملاحة الجوية والبرية .

ومن نماذج قواعد البيانات التي تتيح خرائط مواصلات على الويب:-

قاعدة بيانات Metrokc(٤)

✓ **الخرائط التاريخية** : ترتبط هذه الخرائط بالأحداث والمواقف التاريخية، والاستخدام الرئيس لها يكون في الدراسات التاريخية. كما أن لهذه الخرائط أهمية إثبات الحقائق والمواقف التاريخية، حيث تأخذ مكانًا متقدمًا في أنواع مصادر المعلومات التاريخية .

ومن نماذج قواعد البيانات التي تتيح خرائط تاريخية على الويب:-

قاعدة بيانات Jewishhistory(٥)

---

(١) http://www.torro.org.uk/photos.htm

(٢) http://www.visoo.com

(٣) http://www.metrokc.gov/gis/

(٤) http://jewishhistory.huji.ac.il/links/maps.htm

(٥) http://www.lib.utexas.edu/Libs/PCL/Map_collection/historical/history_main.html

قاعدة بيانات Lib.utexas (١)

قاعدة بيانات lcweb2.loc (٢)

٢/٢/٣/٢ صور الأشعة

أصبحت أجهزة التشخيص التليفزيونية والأجهزة المستخدمة في التحليل الطبي تلقي بالآلاف من اللوحات الطبية التي تدور حول موضوع يتعلق بمرض أو جزء من أجزاء الجسم، ويطلق على هذه اللوحات الطبية اسم صور الأشعة. والشق التنظيمي لهذا النوع من المادة يتم في أغلب حالاته اعتمادًا على التنظيم المستخدم لملفات المرضى أصحاب هذه التحاليل الطبية .

ومن نماذج قواعد البيانات الطبية التي تتيح صور الأشعة على الويب:-

قاعدة بيانات Medicine.ucsd (٣)

قاعدة بيانات .Library.uthscsa (٤)

٣/٢/٣/٢ اللوحات الفنية

وهي مجموعة الرسومات التي تفرزها عقول الفنانين على هيئة إبداع يريد الفنان إيصاله في شكل فكرة إلى ذهن المتلقي مستخدما العرض المبدع وتجانس الألوان. وتحتل هذه اللوحات مكانًا كبيرًا بين مجموعات الصور التي تضمها محركات بحث الصور في قواعد بياناتها بصفة عامة؛ إلا أن اللوحات الفنية لها واقع خاص عند المهتمين بالأعمال الفنية وخاصة الأعمال القديمة، حيث تأخذ مسمى آخر هو المعارض الفنية،كما تأخذ طريقة أخرى في العرض تنبني أساسا على التصفح الذي يلائم أكثر في طريقة التجوال في المتاحف الفنية. وأكثر هذه المتاحف أو المعارض تكون مقتصرة على أعمال فنان واحد أو فترة زمنية محددة.

ومن نماذج قواعد البيانات التي تتيح اللوحات الفنية على الويب:-

قاعدة بيانات Iconbank (٥)

---

(١) http://cweb2.loc.gov/ammem/gmdhtml/gmdhome.html

(٢) http://nedicine.ucsd.edu/clinicalimg/

(٣) http://www.library.uthscsa.edu/

(٤) http://conbank.com/gx/index.htm

(٥) http://www.freeimages.co.uk

قاعدة بيانات Freeimages.[1]

٤/٢/٣/٢ الرسوم الكاريكاتيرية

يتسم هذا النوع من المواد المصورة بالفكرة التي يريد أن يوصلها الفنان إلى القارئ من خلال خطوط وبعض الكلمات المصاحبة لها والتي في الغالب تكون نقدًا لموقف من المواقف ولكن بشكل يتسم بالفكاهة. وهناك ثمة فارق بين (الكاريكاتير) وبين (الكارتون)؛ حيث يهدف الأول إلى إبراز صفات وشخصيات بشكل مبالغ بأسلوب أدبي أو فني. أما الكارتون فهو رسوم هزلية مبسطة وتكون بطريقة رمزية تمثل في الغالب موقفًا أو حدثًا لشخص معين.

ومن نماذج قواعد البيانات التي تتيح الرسوم الكاريكاتيرية على الويب:-

قاعدة بيانات Fotosearch[2]

قاعدة بيانات Google[3]

قاعدة بيانات Cartoonbank[4]

٥/٢/٣/٢ الشرائح

تتكون الشرائح من صور أو لقطات ملونة أو بيضاء وسوداء. وتثبت الشرائح في أطر سميكة، حيث يتم عرضها باستخدام جهاز عرض الشرائح. وحري بالذكر أن الشرائح باعتبارها وسيط معلومات لم توجد لها مكتبات قبل عام ١٨٨٠. وتطورت من الشرائح الزجاجية إلى شرائح الأفلام ٢×٢ بوصة. ويمكن أن تكون هذه الشرائح مستقلة في المعلومات تعمل كإطارات منفصلة، أو أن تكون لقطات متتابعة عن موضوع محدد. وتختلف الشرائح عن الشفافيات فالأخيرة عبارة عن أفرخ من مادة شفافة، غالبًا ما تكون من البلاستيك. وتحمل هذه الشفافيات معلومات قد تكون صورًا أو أشكالا أو رسومات تدل على حقائق معينة. ومن هذه الشفافيات نوعان هما الشكل الملفوف، والشكل المسطح. وهناك أيضا الفيلمات أو الشرائح الفيلمية وهي مجموعة اللقطات المتصلة والتي قد ترقم

---

(١) www.fotosearch.com

(٢) www.google.com

(٣) http://www.cartoonbank.com/

(٤) شعبان عبد العزيز خليفة. مرجع سابق. ص ٢٣-٢٥.

أولا ترقم. ويستخدم لعرض هذه الفيلمات جهاز عرض يدوي أو آلي. وتقف هذه الفيلمات في منطقة وسط بين اللقطات الفيلمية الثابتة والأفلام المتحركة؛ حيث إنها عبارة عن لقطات كان يمكن أن تكون منفصلة لكن تم إبقاؤها على اتصالها معًا.[١]

ومن نماذج قواعد البيانات التي تتيح الشرائح على الويب:-

قاعدة بيانات Worldbank[٢]

٦/٢/٣/٢ الطوابع

تعد هواية جمع الطوابع هي الأساس في جعل الكثير من قواعد الصور تقتني مجموعات من الطوابع لخدمة هذا الهدف والأهداف الثقافية أيضا. ومن المنظور المعلوماتي لهذا النوع من المادة المصورة، فإنها تحتوي على معلومات ذات قيمة تعريفية بالأحداث والمواقف المهمة التي مرت بها كل دولة؛ حيث لا تصمم الطوابع إلا بواقع أحداث أثرت في حياة الدولة .

ومن نماذج قواعد البيانات التي تتيح طوابع على الويب

قاعدة بيانات Alltheweb.com[٣]

قاعدة بيانات Google[٤]

قاعدة بيانات yahoo[٥]

٧/٢/٣/٢ الصور الفوتوغرافية

تستخدم الصور في التعبير عن الأشياء والأحداث والشخصيات بشكلها الذي حدثت به ساعة التقاط الصورة، وهذا التعبير يتم من خلال نقل شكل الحقائق والأشياء على ورق سالب باستخدام مصدر ضوئي. وقد كانت أول آلة تصوير تم استخدامها في عام ١٨٨٨م، وذلك عن طريق شركة كوداك باستخدام الأفلام الملفوفة. ثم تطورت تقنيات التصوير فأصبح هناك تصوير بالأقمار الصناعية، وأخرى ميكروسكوبية، واستخدام آلات التصوير الرقمية مع باقي أنواع التصوير

---

(١) http://www.worldbank.org/html/extdr/emps/photo.htm

(٢) http://www.luxcentral.com/stamps/LuxStampsDB.html

(٣) wwww.google.com

(٤) www.yahoo.com

(٥) www.corbis.com

الأخرى. أما عن سبل إتاحة الصور، فهي إما ورقية أو فيلمية أو رقمية. وتعد الأخيرة هي موضوع هذه الرسالة وسوف نتعرض لها بالتفصيل.

ومن نماذج قواعد البيانات التي تتيح صورًا فوتوغرافية على الويب:-

قاعدة بيانات Corbis(١)

قاعدة بيانات Freefoto(٢)

قاعدة بيانات Ditto(٣)

٤/٢ المادة المصورة في بيئة الويب

تمثل عملية التكشيف البداية لأي حديث عن استرجاع المادة المصورة، ذلك لأن الاسترجاع إنما يكون وفقًا للتنظيم والمعالجة الفنية للمادة، والتي تعتمد بدورها على الوصف المادي والتحليل الموضوعي لها. وكما هو معروف لدى كل اختصاصيي المعلومات، فإن عناصر الوصف المادي والتحليل الموضوعي هي المحور الرئيس للمادة المصورة في البيئة التقليدية، ولا يختلف الأمر عن ذلك في بيئة الويب باستثناء عدد من العناصر التي أضيفت لأغراض الاسترجاع. وتتنوع هذه العناصر من مادة مصورة إلى أخرى، وحيث إن الحديث عن كل أنواع المواد المصورة يتعدى حدود هذه الرسالة، فإننا سنقصر الحديث هنا على الصور الفوتوغرافية فقط باعتبارها موضوع هذه الدراسة.

١/٤/٢ الصور الفوتوغرافية في بيئة الويب

إن الصور الفوتوغرافية الرقمية على الويب تشهد نموًا متزايدًا في أعداد مواقعها وحجم قواعد بياناتها، وقد ظهرت أنواع مختلفة من قواعد بيانات الصور الفوتوغرافية، دفعت القائمين على المواقع إلى التخصص. وتتعدد المجالات التي تضمها هذه الصور؛ إلا أنه لأغراض هذه الدراسة يمكن تقسيمها إلى ثلاثة أنواع أساسية هي:-

---

(١) www.freefoto.com

(٢) www.ditto.com

(٣) http://www.astronomy.ca/images/

- **الصور الفوتوغرافية العلمية** : هي الصور التي تختص بصور العلوم الطبيعية والبيولوجية والزراعية والطبية وما ينشق عن هذه العلوم. وتبرز أهمية الصور في هذه العلوم؛ لأن الوصف بالكلمات قد يكون غير مناسب في بعض المواضع بل يكون مستحيلا في الكثير منها، فصور الأرض والمريخ، وصور البراكين والزلازل، وصور البكتيريا والفيروسات، وصور أمراض النبات، وصور الأورام الخبيثة وغيرها تكون الصورة فيها هي الأصل ويصاحبها بعض الكلمات للربط بين اللقطة والمسمى العلمي للمحتوى .

ومن نماذج قواعد البيانات التي تتيح صورًا فوتوغرافية علمية على الويب

قاعدة بيانات في مجال الفضاء Astronomy[١]

قاعدة بيانات في مجال الجيولوجيا Wordweb[٢]

- **الصور الفوتوغرافية الإعلامية** : وهي مجموعة الصور التي تعبر عن الأحداث الجارية، إضافة إلى كونها الركن الثاني للإعلام المقروء بجانب النص. وظهرت الصور الرقمية الإعلامية على الويب بإتاحة الصحف الرقمية على الويب .

ومن نماذج قواعد البيانات التي تتيح صورًا فوتوغرافية إعلامية على الويب :-

قاعدة بيانات 911digitalarchive[٣]

قاعدة بيانات كشاف واشنطن بوست Photostore[٤]

- **الصور الفوتوغرافية العامة** : وهي مجموعة الصور التي لا تحوي أشكالا علمية أو ظواهر بحثية تنتمي إلى مجال موضوعي معين، وإنما هي مجموعة من الرموز وعناصر من الطبيعة العامة أو أشخاص أو أحداث يراد من خلالها إيصال معلومات محددة. وتعد الصور الرقمية الطبيعية هي لب هذه الدراسة ومحورها الأساس بالنظر إلى خصائص ومحتوى هذا النوع وكيفية التعامل معه

---

(١) http://wordweb.com/docs/gallery/geo/

(٢) http://911digitalarchive.org/

(٣) http://www.photostore.com/psx/index.html

(٤) www.corbis.com

بما يوفر أفضل استرجاع ممكن. ومن نماذج قواعد البيانات التي تتيح صورًا فوتوغرافية عامة على الويب:-

قاعدة بيانات Corbis[(١)]

قاعدة بيانات Freefoto[(٢)]

قاعدة بيانات Ditto[(٣)]

وتتنوع قواعد بيانات الصور الفوتوغرافية من حيث العدد والمحتوى على الويب. فلقد لمع فضاء الويب بقواعد بيانات الصور الطبيعية مع إطلاق مسمى آخر هو قواعد بيانات الصور العامة، ذلك لأن قواعد بيانات الصور العلمية اعتمدت في نشأتها على الجمعيات والجامعات الأقسام العلمية، على الجانب الآخر من قواعد بيانات الصور فإن محركات بحث الصور لم يظهر منها حتى الآن النوع المتخصص في بحث الصور العلمية، وإنما تحمل كل أنواع محركات بحث الصور مصطلحًا (عاما). والاختلاف بين قاعدة بيانات الصور ومحرك بحث الصور هو اختلاف بيّن يرجع إلى الاختلاف في طبيعة اقتناء الصور بجانب طبيعة تكشيف ملفات الصور. ويثبت ذلك أن الصور العلمية لا يوجد لها محركات بحث الصور بحسب ما يقتضيه مصطلح محرك البحث، وإنما توجد لها قواعد بيانات تتخصص بعرض صور مجال موضوعي معين، بينما تسكن الصور الطبيعية قلب كل من قواعد بيانات الويب ومحركات بحثها، بحسب ما يقتضي مفهوم كل من المصطلحين.

ومن ناحية أخرى فقد تكون قواعد البيانات متخصصة بالشكل الذي فصلناه، وقد يجتمع نوعان منها أو الثلاثة معًا في قاعدة بيانات واحدة. فوفقا لما رآه الباحث من هذه القواعد، فإن قواعد بيانات الصور العلمية لا يمكن أن تحوي صورًا طبيعية أو عامة.. بينما يمكن لقواعد الصور العامة أن تحوي بين مجموعاتها صورًا علمية.

وتأخذ الصور العامة نفس خصائص الوصف الببليوجرافي التي يمكن العمل بها على كل من الصور العلمية والإعلامية، إضافة إلى استخدام كل عناصر وحقول

---

(١) www.freefoto.com

(٢) www.ditto.com

(٣) A Metadata Framework to Support the Digital Resource Management. Texas university, 2003, visited at 10/10/2003, http://texashistory.unt.edu/guides/Draft-UNTL-Metadata-Guideline.htm

التسجيلة الببليوجرافية إلا ما يندرج أسفل نوع بذاته، وهذا يرجع إلى توافق الأنواع الثلاثة في خصائص الصورة كمادة مرئية، والتشابه في ملفات الصور الرقمية بحسب الشكل Format الذي ستكون عليه صور قاعدة البيانات. وترجع إرهاصات معايير ومواصفات الصور إلى الاعتماد على نوع الصور العامة؛ مثل ما كان عليه معيار دبلن كور، وهو ما سيأتي الحديث عنه في معالجة الصور الرقمية.

٢/١/١/٤ الوصف المادي للصور الفوتوغرافية على الويب

تركز الوصف المادي للصور الفوتوغرافية في الويب على عناصر ثلاثة أساسية هي: المصور، والنوع، والشكل الفني ...الخ. وهو في ذلك يتشابه إلى حد كبير للوصف المادي للصور الفوتوغرافية في البيئة التقليدية. ثم أضافت بعض قواعد البيانات عناصر أخرى هي اللون والشكل shape في عمليات المضاهاة واسترجاع الصور.

وقد حاولت بعض الهيئات صياغة نماذج لما يمكن أن تكون عليه حقول الوصف داخل ملفات الصور. واعتمدت جامعة تكساس Texas University نموذجا لحقول الأشكال المختلفة من المادة داخل الويب. وخصص النموذج مجموعة من الحقول الجديدة لوصف ملفات الصور داخل بيئة الويب، وكانت كما يلي:[(١)]

الجدول رقم (٢-١) عناصر جامعة تكساس لوصف الصورة الرقمية

وصف ملف الصورة	
شكل إصدارة ملف الصورة	TIFF v 4.0
كثافة العرض داخل الصورة RESOLUTION	600 dpi; 300 dpi, 1500 d/cm
أبعاد الصورة	4096 x 6144 pixels
كثافة ألوان الصورة	1-bit; 8-bit grayscale; 24-bit color
توزيع الألوان IMAGE COLOR SPACE	CMYK; RGB
ضغط الملفات	zip file, CCIT 4
أخرى	

(١) Martinez, Aanne. GIF vs. JPG vs. PNG. Mc GrawHill companies, 2001 inc, citd (20/7/2002), (http://www.cheapwebrich's.com/articles/gif_jpg_png.html)

وعلى الرغم من هذه الجهود، فإن المحاولات الجارية للوصول إلى تسجيلة معيارية للصورة مازالت قيد البحث، وذلك لارتباطها بالجهات المهنية التي تنتج الصور. فعلى سبيل المثال توجد ثلاثة أشكال فنية formats أساسية للصور الرقمية هي:(١)

- **شكل GIF**: وهو اختصار لـ (graphic interchange format) وهذا النوع من خصائصه الخط البسيط في الرسم وقلة الألوان في الصورة، كما أنه يدعم الرسوم الحيوية. وقد تم إنتاج هذا النوع من جانب شركة CompuServe، وذلك في منتصف الثمانينات، وكان هذا لأجل أن يتم سرعة نقل الرسوم الكبيرة عن طريق تصغير حجمها بالضغط لملفات هذه الرسوم. وموضع استخدام هذا النوع في عمل الأيقونات icons وهي الرسومات البسيطة واللقطات الصغيرة للصور. ويوجد داخل هذا الشكل نوعان هما 87a, 89a، أما النوع 89a فهو يستخدم بصفة خاصة في أنواع الملفات التي يمكن إظهار لون مخصص من ألوان الصورة وعرضه بغض النظر عن إظهار أي من الألوان الأخرى، سواء كانت الحدود الأربعة للصورة أو الألوان الداخلية للصورة. ويرى الباحث أنه يمكن استخدام هذا النوع في نظم الاسترجاع التي لا تعتمد على النص المصاحب للصورة، بل تسترجع بمضاهاة الألوان داخل الصورة.

- **شكل JPEG**: وهو اختصار (Joint Photographs Expert Group) وهذا النوع يدعم الصور ذات الألوان الكثيرة، كما أنه يدعم اختزان وعرض الصور الفوتوغرافية أكثر من الأنواع الأخرى. وينطق هذا الاختصار حسب المقطع الصوتي (jay-peg). ويعتبر هذا النوع هو الأكثر استخدامًا على الويب. ويعود ذلك إلى أنه يمكن ضغط ملفات الصور وتصغيرها مما يؤدي إلى كفاية في عملية الحفظ والاسترجاع، ويمكن ضغط هذا النوع من الملفات بنسبة ١:٥ إلى ١٥:١.

- **شكل PNG**: ويعني به مصطلح Portable Network Graphics. وهذا النوع هو الأحدث بين كل الأنواع السابقة ولذلك لا يدعم استخدامه الكثير من محركات البحث؛ إلا أنه نتيجة لمميزاته سوف يصبح الأكثر شيوعًا على الويب فيما

---

(١) Martinez, anne GIF vs. JPG vs. PNG. mcgrawhill companies, inc,2001, citd (20/7/2002), (http://www.cheapwebrich's.com/articles/gif_jpg_png.html)

بعد. وقد صمم هذا النوع تحديدا ليكون بديلا عن استخدام شكل GIF. ومازال هذا النوع ينتظر الانتشار بين مواقع الويب المختلفة.

وهذه الأنواع هي الأكثر استخدامًا في محركات البحث؛ إلا أن هناك أنواعًا أخرى كثيرة يمكن العمل بها في محركات البحث لكنها أقل استخدامًا وهي:[١]

- **شكل BMP**: وهذا الشكل يعد من الأشكال الأقل استخدامًا على الإنترنت ذلك لأنه قد تم إنتاجه للعمل في بيئة نظام التشغيل ويندوز Microsoft windows operating system. وحجم هذا الملف كبير لذا لا يتم التعامل في محركات بحث الإنترنت. كما أن هذا الشكل لا يدعم التعامل مع النظم الأخرى.
- **شكل TIFF**: وهو اختصار Tag Image File Format، وقد تم إنتاج هذا الشكل في منتصف الثمانينات من قبل شركة متخصصة في إنتاج صور الويب وأسطح المكتب Desk tops. وأنتج هذا الشكل خصيصا للحفاظ على الصور التي يتم نقلها في ملفات ذات جودة عالية. ويعمل هذا الشكل على أكثر من نظام تشغيل مثل ويندوز windows، أجهزة Macintosh، يونيكس UNIX. ويمكن لهذا الشكل أن يكون في وضع مضغوط أو غير مضغوط.
- **شكل PICT**: وهذا الشكل قد تم إنتاجه أيضا في منتصف الثمانينات، على أن يستخدم لملفات الرسومات الضخمة مثل ملفات الفيديو والرسوم الحيوية، وعلى ذلك فهو خلاف الاتجاه الذي يسير عليه شكل JPEG كشكل للملفات الخفيفة.

وتستخدم هذه الأنواع بحسب مميزات كل نوع منها، وأهمية هذه الأنواع في أن نظم استرجاع الصور تعتمد في أحيان كثيرة على الشكل وخصائصه كما في الملفات الرقمية. هذا إلى جانب مجموعة من الأشكال التي تحدثت عنها جمعية الخدمة الفنية الاستشارية للصور[٢] TASI والتي تتلخص في الآتي :-

---

(١) File Formats. TASI Technical Advisory Service for Images, 2003, VISITED AT 24/4/2003 , http://www.tasi.ac.uk/advice/creating/fformat.html

(٢) Eriksson Yvone. How to make tactile pictures understandable to blind reader, international federation of library associations and institutions(IFLA), 1999, visited (1/12/2001), http://www.tasi.ac.uk/advice/creating/fformat.html

شكل PCX: وهو اختصار Windows PaintBrush، ولا يختلف كثيرًا عن أشكال TIF and PNG؛ إلا أن هذا النوع لا يستخدم في بيئة تشغيل الويندوز، وهو على درجة من الفائدة في الأداء مع أجهزة الماكينتوش .Macintosh

شكل TGA: يعتبر هذا الشكل أكثر أشكال ملفات الصور شيوعًا من حيث جودة الكثافة العالية للصور ويمكن الحصول على هذا الشكل من خلال أجهزة إدخال الصور الرقمية كما أنه يدعم درجات اللون المختلفة مثل ٨ ١٦ ٢٤ ٣٥ بت BIT.

شكل EPS: وهو اختصار Encapsulated PostScript وقد أنتجته شركة Adobe Systems Inc. ويمكن لهذا الشكل العمل على أكثر من نظام تشغيل فضلا عن إمكانية تحويل هذه الملفات إلى الأشكال الأخرى. ويستخدم شكل EPS للعمل مع الملفات المراد طباعتها.

شكل SVG: وهو اختصارScalable Vector Graphics ويستخدم هذا الشكل بشكل معياري مع لغة XML، (Extensible Markup Language) وهي النموذج المتقدم للغات الترميز المعيارية ويستطيع هذا الشكل أن يضم نصًا مصاحبًا للرسومات، كما يمكن إدخال التحرير بعيدًا عن الصورة، وهذا خلاف ما يحدث في الأنواع الأخرى حيث يظهر النص داخل إطار الصورة.

شكل CGM: وهو اختصار Computer Graphics Metafile، وقد تم تصميم هذا الشكل بواسطة المعهد القومي الأمريكي للتقييس ANSI لكي يتم وضع بناء أولي لتبادل مواد الرسومات الرقمية.

شكل WMF: وهو اختصار Windows Metafile ويقوم هذا الشكل على احتواء أشكال المخططات الثانية والرسومات كي يتم رفعها إلى درجة الكثافة المطلوبة على شاشة العرض.

شكل RIFF: وهو اختصار Resource Interchange File Format وهو شكل الملف متعدد المصادر، ويتكون هذا الشكل من مجموعة بنى متداخلة خلاف ما تدعم الوسائط المتعددة. وقد خرج هذا الشكل بواسطة شركتي Microsoft and IBM.

شكل DIB: وهو اختصار Device Independent Bitmap وتعمل ملفات هذا الشكل داخل بيئة التشغيل windows. ويمكن عرض هذه الملفات من خلال أجهزة عرض مختلفة. ويشابه هذا النوع شكل DMP لعرض الصور والذي يدعم فتح

ملفات الصور في أكثر من حزمة برامج. ويعد هذا النوع شائع الاستخدام في حزم البرامج التي تتطلب فتح ملفات الصور في أكثر من نوع من البرامج.

- **شكل FlashPix**: تم إنتاجه بواسطة شركة Kodak وقد أطلق عليه الشكل المفتوح من ملفات الصور. ويساعد تصميم هذا الشكل على التفاعل مع التطبيقات المختلفة بشكل متوازن؛ حيث يمكنه خفض الكثافة resolution في حالة استخدام شاشات العرض، أو رفع الكثافة أثناء الحاجة إلى الطباعة. كما أنه يتسم بعدة ملامح مختلفة تلائم حاجة السوق الذي تتعدد فيه استخدامات الصور؛ مثل حقول البيانات، وحق النشر، وتنظيم وتهيئة الألوان.
- **شكل SPIFF/SPF**: وهو اختصار Still Picture Interchange File Format، وهو بمثابة الشكل الرسمي لشكل JPEG. ومازال استخدام هذا الشكل قليلًا نسبيًا نظرًا لحداثة خروجه؛ إلا أنه من المتوقع انتشاره نظرًا لأنه يدعم الكثير من تطبيقات ملفات الصور.

وتظهر أهمية هذا التنوع في Formats عند استخدام عناصر بناء لغة الترميز المعيارية للنص الفائق HTML. ولأن محركات البحث كانت الخطوة التالية بعد بناء صفحات الويب، فقد بني عمل هذه المحركات في الاقتناء والتكشيف والاسترجاع على بنية ملفات الـHTM، وليست ملفات الصور ببعيدة عن هذا التعامل، بل أنها تأخذ كل خصائص ملفات النص الفائق والوسائط الفائقة Hyper Text and Hyper Media في الإضافة والاسترجاع. ولا يختلف الأمر بين أنواع الملفات إلا بوجود النص.

٢/١/٤/٢ التحليل الموضوعي للصور الفوتوغرافية على الويب

يحتل التحليل الموضوعي درجة أكثر من الأهمية عن نظيره المادي في الاسترجاع على الويب؛ ذلك لأن البحث دائما في الويب يكون حسب الحاجة لموضوع معين. ويعتمد التحليل الموضوعي للصور على الويب على عنصرين أساسين الأول: طبيعة الصورة، والثاني: السياسة التي سيتبعها المكشف لهذه الصور.

فبالنسبة للعنصر الأول، فإن استخدام الكلمات المفتاحية للصورة العلمية يختلف عنها في الإعلامية أو في الطبيعية. ففي الصور العلمية، لا يكون هناك مجال كبير للاجتهاد بالنسبة للمكشف، حيث يجب عليه استخدام المصطلحات العلمية الواصفة للصورة، فالصور العلمية دائما ما تحوي أشكالا أو عناصر ذات

دلالة ثابتة، وعن طريق ذلك فإنه من اليسير إيجاد التوافق بين المكشف والمستفيد من نظام الاسترجاع. ويمكن لعملية التكشيف والاسترجاع أن تصل إلى درجة التحقيق العالية من خلال الاعتماد على مكنز لهذا الموضوع أو العلم الذي بصدده عملية التكشيف. ويقوم هذا المكنز بإيجاد التقنين واستخدامه كأداة للمكشف أثناء التنظيم واستخدامه كقائمة للاستناد للباحث أثناء الاسترجاع. وعلى ذلك فإن الدلالة التي تكون عليها مصطلحات هذه العلوم سوف تسهم في الوصول إلى الدرجة العالية في التحقيق وفي الوقت نفسه تحمي المستفيد من نسب الشوشرة (النتائج غير الصحيحة) التي تؤثر على النتائج المسترجعة.

أما في الصورة الإعلامية، فإن هناك مساحتين للمكشف الأولى: الحاكمة وهي تلك التي تصف عناصر الصورة مثل الشخصيات أو الأماكن أو الأحداث التي تظهر في الصورة. وهذه العناصر يمكن أن تحكمها قوائم استناد لتوحيد المسميات الخاصة بالشخصيات، والمناطق الجغرافية، والأحداث. أما الثانية: وهي غير الحاكمة ويعبر فيها المكشف عن خلفية الصورة أو الانفعالات التي تطرحها الصورة مثل الضحك، والعبوس، والوقوف، والجلوس وغيرها، وهو ما يعرف باسم الصورة الذهنية. وبالتأكيد فإن وصف هذه الانفعالات يعتمد بشكل رئيس على الانطباعات التي تتركها الصورة في نفس المتلقي ومن ثم تنعكس على طريقة توصيفها، وهي بلا شك أكثر عرضة للاختلاف بين المكشف والباحث وتولد درجة غير قليلة من الشوشرة.

أما في الصور العامة، فهناك ثلاث مساحات للمكشف، الأولى: الحاكمة وهي التي تصف العناصر الرئيسة في الصورة مثل الجبال والأنهار والطيور وغيرها وهذه يسهل توصيفها، والثانية: الذهنية وهي التي تصف الخلفية للصورة وما تتركه من انطباعات وهذه أكثر عرضة للاختلاف بين المكشف والباحث كما هو الحال في الصور الإعلامية، والثالثة: الضمنية وهي التي تخرج عن دائرة الوصف الموضوعي إلى المعنوي، كأن يوضع لفظ (جمال) ككلمة مفتاحية للتعبير عن صورة : وردة متفتحة". ومما لا شك فيه أن هذه المساحة الثالثة هي أكثرها صعوبة على الإطلاق نظرًا لعدم وجود قواعد يمكن الرجوع إليها بشأن تلك المعاني الضمنية.

علاوة على ذلك فإن الحديث عن استرجاع الصور الفوتوغرافية في بيئة الويب، أمرغير يسير، وذلك لأنه يتضمن تغيير لغة التعبير عن المحتوى من وسط إلى آخر؛ أي من وسط الصورة إلى وسط النص. وإضافة إلى ذلك فإن اختيار

أفضل الكلمات للتعبير عن المحتوى أمر يعود إلى وجهة نظر المكشف، وقد توصف الصورة بأكثر من وصف للمحتوى ولا خطأ في أي منها. وبذلك فان الأمر يحتاج إلى الوصف بدرجة أكبر من التحليل ويتم ذلك على أربع مراحل؛ أولا: أن يتم وصف الصورة بشكل عام، ثانيا: أن يتم تحديد كل عنصر من عناصر الصورة للاقتراب أكثر من محتوى الصورة الموضوعي، ثالثا: أن يتم وصف الصورة بشكل مفصل يجعل ملامح الموضوع أكثر وضوحا للقارئ، أما المستوى الأخير فهو إعطاء وصف ملخص عن الصورة بشكل جامع لكل ما يمكن الحديث عنه.(١)

أما بالنسبة للسياسات فيواجه المكشف مجموعة من الأسئلة التي يجب أن يعرف إجاباتها سلفا حتى يحدث الاتساق والاطراد في التكشيف بمعنى هل تكون المصطلحات المفردة أو الواصفات المستقلة هي الأفضل أم النص السردي والمستخلص هي الأكثر فاعلية ؟ ولقد لاحظ الباحث في وصف محركات البحث لمجموعاتها من الصور. أنها تكون باللغة الطبيعية أقرب في حالتها إلى المستخلص الذي يسرد موضوع ما وهو ما يوافق طبيعة الصورة؛ ذلك لأننا عندما نقول إن الصورة تفوق آلاف الكلمات، فإن هذا الكم من الكلمات لموضوع من الصعب التعبير عنه بكلمات قلائل في شكل منفرد مثل الواصفات أو الكلمات الكشفية.(٢)

وعلى ما سبق يمكن القول إن استخدام حقول البيانات الواصفة داخل ملفات الصور الرقمية يعني إضافة بيانات معبرة عن شكل الصورة ومحتواها سواء في صفحة الويب الأصلية أو داخل قاعدة محركات بحث الصور أو داخل منفذ المستخدم النهائي عند تحميل الصورة أو إرسالها emailed or downloaded. وما تحتاجه هذه الدراسة هو التنبيه على البيانات الملازمة للصورة والمعبرة عنها حيث تؤثر هذه البيانات في تزويد وتكشيف واستدعاء الصورة داخل محركات بحث الصور التي تعتمد كلية على الحقول المصاحبة للصور الرقمية؛ مثل حقل الكلمات المفتاحية، والعنوان، نوع وشكل الصورة، ووصف الصورة ... إلخ.

---

(١) Notess, Gerg R. op cit. p3.

(٢) Image Tagging: Embedding Metadata In Image Files.
TASI Technical Advisory Service for Images, 2003, visited at 28/4/2003,
http://www.tasi.ac.uk/advice/delivering/faqtagging.html

ولقد خرجت الأنواع السابقة من ملفات الصور الرقمية بالشكل الذي تفارق فيه الصورة حقول البيانات - المحدودة - المصاحبة لها عند نسخ أو تحميل هذه الصورة، مما يؤثر بدوره على تحويل ماهية الصورة واختلاف البيانات المعبرة عنها. ويمكن تمثيل المعنى السابق بمجموعة الصور داخل حافظة واحدة وبجانب كل صورة يتم وضع البيانات المعبرة عنها مثل تاريخ إنتاج الصورة أو المصور أو المكان، وعلى ذلك يتضح محتوى الصورة أينما ذهبت. وتعد قواعد بيانات الصور الرقمية الآنية على هذا النحو من العمل، حين تضع حقول البيانات الواصفة فقط لتكشيف واستدعاء الصور الرقمية داخل قاعدة البيانات فقط.[١]

ويمكن إرجاع الحاجة لوصف ملفات الصور بحقول البيانات Tagging إلى أن مختلف البرامج التي تعمل على ملفات الصور لا تتوافق في فتح ملفات الصور مما ألزم مصممي هذه الملفات إلى إيجاد صيغة توافق مختلف البرامج والتطبيقات الخاصة بالصور الرقمية.

ويختلف حجم حقول البيانات المصاحبة لملفات الصور بحسب اتجاه استخدام قاعدة البيانات، بحيث إذا كانت قاعدة البيانات لا يحتاج مستخدموها إلى الحقول الكثيرة؛ فان الملفات التي تأخذ متوسط مساحة أقل هي الأكثر كفاية عن الملفات الأخرى. أما إذا كانت قاعدة البيانات تخدم مستفيدين ذوي طابع خاص فانه من اللازم توافر حجم أكبر من الحقول ذات الخصائص الأكثر تفصيلا عن الصور، على أنه في كل الحالات فانه لابد من وجود عدد من الحقول المعيارية يعبر عن ماهية الصورة الشكلية والموضوعية.

٢/٤/٢ الصيغ المعيارية لوصف الصور الرقمية

لقد كان لتباين الاختلافات في الوصف المادي للصور، ومشكلات التحليل الموضوعي لها أن اتجهت الكثير من المنظمات المتخصصة في مجال المعلومات وغير المتخصصة على أن تصنع معايير ومواصفات ثابتة يمكن من خلالها التحكم الببليوجرافي والمعلوماتي في أوعية معلومات الصور. وهو ما يعتبر خطوات واسعة في طريق الاهتمام بنظم استرجاع الصور وتحقيق أعلى معدلات من الدقة والكفاية في استدعاء معلومات الصور .

---

(١) Graham, Margaret E. THE DESCRIPTION AND INDEXING OF IMAGES:

ولقد عمل المسح الذي أجراه معهد أبحاث البيانات المصورة Institute for IMAGE DATA Research[١] على إظهار الاستخدام الحقيقي لمعايير وصف الصور داخل المكتبات التقليدية. وقد جاءت النتائج في صالح استخدام معايير وصف ببليوجرافي يتم إعدادها داخل المكتبة والملائمة لطبيعة استخدام المكتبة والمستفيدين لمجموعاتها من الصور الفوتوغرافية؛ حيث كانت نسبة وصف الصور الفوتوغرافية بمعايير وصف داخلية هي ٧١٪ ويلي ذلك قواعد الفهرسة الأنجلو أمريكية AACR2 بنسبة ٢٥٪ ثم قواعد الفهرسة المقروءة آليا MARC بنسبة ١٧٪ ويوضح الجدول رقم (٢-٢) أنواع ونسب استخدام قواعد الوصف الببليوجرافي للصور داخل المكتبات موضحًا معايير الأنواع الأخرى من المواد غير النصية:-

الجدول رقم (٢-٢)

يوضح مراتب استخدام معايير الوصف الببليوجرافي للصور داخل المكتبات

Other	In-house rules	SPECTRUM	VRA (Visual Resources Association)	Metadata - Dublin Core	MARC	AACR	No.*	Cataloguing standard: Type of Material:
٪٩	٪٧٢	٪٢	٪٤	٪٢	٪٩	٪٢٥	٥٣	الشرائح (35 mm)
٪٢	٪٢٠	٪٢	-	-	٪٢٥	٪٦٤	٤٤	الفيديو
٪٤	٪٧١	٪٤	٪٤	-	٪١٧	٪٢٥	٢٤	الصور الفوتوغرافية (positives)
٪٥	٪٦٨	٪٥	٪٥	-	٪١٦	٪٢٦	١٩	الصور الفوتوغرافية (negatives)
٪٦	٪٤٤	٪٦	٪٦	-	٪٦	٪١٩	١٦	الملصقات
٪٧	٪٦٤	٪٧	-	-	٪٢١	٪٣٦	١٤	المطبوعات
٪٨	٪٦٩	٪٨	-	-	٪١٥	٪٣١	١٣	اللوحات
٪٨	٪٦٧	٪٨	٪٨	-	٪١٧	٪٣٣	١٢	الرسومات
٪١٨	٪٤٥	٪٩	٪٩	-	٪٩	٪١٨	١١	الشفافيات ("5"x4)
٪٢٥	٪٥٠	٪١٣	-	-	٪١٣	٪٣٨	٨	لوحات القماش
٪١٤	٪٥٧	٪١٤	-	-	٪١٤	٪٢٩	٧	الأفلام
٪٢٠	٪٤٠	-	-	-	-	٪٢٠	٥	المنتجات الفنية

(١) Report of a survey of ARLIS members, Institute for IMAGE DATA Research1998/99, visited at 6/10/2003, http://www.unn.ac.uk/iidr/ARLIS/.

ويلاحظ الباحث أن أحدًا من أفراد المسح لم يشر إلى استخدام معيار Metadata - Dublin Core في الوصف الببليوجرافي للصور، وذلك إنما يأتي من ندرة في التفاعل بين مجموعات الصور داخل المكتبات التقليدية وبيئة الويب. ومن ثم فقد قام الباحث بمسح للمعايير التي وضعت للصور في البيئة الرقمية من أجل التعرف إلى أكثرها ملاءمة لتلك البيئة.

٢/٤/٢/١ معيار NISO

وضع هذا المعيار منظمة National Information Standards Organization (NISO)، وهي من المنظمات غير الهادفة للربح قام بتأسيسها المعهد القومي الأمريكي للتقييس American National Standards Institute وتعرف اختصارا باسم ANSI عام ١٩٣٩م ثم تغير اسمها إلى الاسم الحالي وذلك بغرض رعاية وصيانة وإنتاج المعايير اللازمة لمجال المعلومات. ومن أهم المعايير التي قدمتها هذه المنظمة هي /ANSI NISO Z39.7 - 1995 Library Statistics, ANSI/NISO 12083 - 1995 (R2002) Electronic Manuscript Preparation and Markup, ANSI/NISO Z39.14 - 1997 (R2002) Guidelines for Abstracts, ANSI/NISO Z39.50 - 1995 Information Retrieval : Application Service Definition & Protocol Specification. وتدور محاور معيار الصور حول عملية الوصف والتعبير عن المحتوى الموضوعي لها، وقد انتهت المسودة الأخيرة منه في عام ٢٠٠٠م[١].

٢/٤/٢/٢ معيار DIG35 Specifications

وضع هذا المعيار هيئة International Imaging Industry Association، وهي هيئة تجمع بين مجموعة من الشركات العاملة في مجال إنتاج الصور الرقمية مثل Canon, Agfa, Kodak... وغيرها. ويحتوي المعيار على مجموعة من الحقول الموحدة المستخدمة لوصف الصور الرقمية ويرمز له بـ I3A لتمثيل اسم الجهة القائمة على بنائه. وكان أول إصدار لهذا المعيار في الثامن عشر من يونيو عام ٢٠٠١م وذلك بإخراج الإصدارة ١,٠، ثم يلي ذلك خروج الإصدارة ١,١ والتي تعمل بالتوافق مع ترميز لغة الترميز المعيارية القابلة للتوسع XML Tags؛ وذلك لما لهذه اللغة من قدرة على توفير البيئة الملائمة لعمل حقول الوصف الميتاداتا وصلاحيتها لاستدعاء

(١) About NISO, niso, 2001, visited at 20/9/2002, http://www.niso.org/about/index.html

الصور الرقمية وعرضها بشكل فردي. ولقد حرصت هيئة الإصدار على بناء هذا المعيار بالشكل الذي يمكن معه مواجهة صعوبات التعامل مع ملفات الصور الرقمية وتنظيمها واسترجاعها، وهي الصعوبات التي أخذت في التزايد مع انطلاق عمل محركات بحث الصور في تنظيم وبحث الصور. كما عملت هيئة إصدار هذا المعيار على صيانته في إصداراته المختلفة ليتوافق مع المعايير الأخرى مثل ISO's JPEG-2000 and MPEG-7، بالإضافة إلى متابعة التحديث في لغة XML ولغات التركيز المعيارية الأخرى[١].

٣/٢/٤/٢ معيار VRA 3

يتناول هذا المعيار the Visual Resoirce Association Core categories v3 (VRA 3) ويبلغ عدد عناصره ١٧ عنصرًا ترتكز على أوعية المعلومات المصورة غير أنها ليست في نفس تعقيد معيار مارك Machne Readable Cataloguing (MARC). وتتعدى عناصر وصف هذا المعيار أوعية الصور الثابتة الرقمية إلى مختف المعلومات المرئية الأخرى مثل فنون الرسم على الجدران وفنون الأشكال الهندسية إلى الصور المطبوعة[٢].

٤/٢/٤/٢ معيار Metadata for Long term Preservation

وضع هذا المعيار The Neworked European Deposit Library (NEDLIB) وعرف باسم Metadata for Long term Preservation، ويتعلق بالعناصر والحقول الجوهرية المستخدمة في وصف المصادر الإلكترونية وخاصة الصور التي يتم حفظها لأمد طويل وجاء المعيار مركزا على ثمانية عناصر رئيسة وثمانية وثلاثين عنصرًا فرعيًا .

ويوازي مصطلح الميتاداتا هنا عملية الوصف الببليوجرافي .اخل المكتبات التي تعمل على وضع حقول لوصف شكل ومحتوى المادة داخل الكتبات. ويعد استخدام مصطلح الميتاداتا مصطلحًا عامًا إذا ما أردنا به الحديث عن البيانات أو

---

(١) DIG35: Metaidaa Standard for Digital Images. XMLCoverPages, 2002, visited at 21/6/2004, http://xml.coverpages.org/dig35.html

(٢) Metadata: Imae Description, Indexing and Cataloguing. tasi, 2004, visited at 21/6/2004, http://www.tasi.ac.uk/advice/delivering/meta.html

الحقول الواصفة المادة داخل بيئة الويب؛ حيث تستخدم حقول الميتاداتا على مستويات أكثر تخصصًا تتجلى في ثلاثة مستويات هي:-[(١)]

أولًا : استخدام حقول الميتاداتا لوصف ملفات html أي الملفات النصية فقط داخل بيئة الويب.

ثانيًا : استخدام الحقول الميتاداتا لوصف الملفات غير النصية المصاحبة لصفحات الويب مثل ملفات الصوت والصورة والفيديو.

ثالثًا : استخدام حقول الميتاداتا لوصف الملفات التي توجد داخل مختلف قواعد البيانات على الويب والتي تلائم استخدام تلك القاعدة دون غيرها كالتي يتم العمل بها داخل محركات البحث أو غيرها.

ولقد ذكر Diann Rusch-Feja أن هناك ثلاثة أشكال تتخذها حقول الميتاداتا عندما ترفق بمصادر المعلومات لبناء الوصف الببليوجرافي وهي كما يلي[(٢)]:-

١ - صياغة حقول البيانات الخلفية كحقل <META Tags> داخل الجزء الأعلى من صفحات اللغة المعيارية HTML.

٢ - صياغة حقول البيانات الخلفية كملف منفصل <META =...> عن صفحة الويب لوصف الملفات التي تصحب صفحات الويب وليست جزءًا منها مثل ملفات الصور الرقمية والصوت والملفات التنفيذية.

٣ - صياغة البيانات الخلفية كجزء مكون لقاعدة البيانات التابعة لأحد خادمات الويب أو لأحد محركات البحث على الويب والتي تتميز باستقلالها في البناء عن غيرها من البيانات داخل قواعد البيانات الأخرى.

٢/٤/٢/٥ معيار دبلن كور Dublin core

تم الاتفاق على إصدار هذا المعيار في عام ١٩٩٥م في مدينة دبلن Dublin في ولاية أوهايو Ohio بين مجموعة من الهيئات المنتشرة في مختلف الدول مثل

---

(١) Rusch-Feja, Diann. Metadata: Standards for Retrieving WWW Documents (and Other Digitized and Non-Digitized Resources), Astronomical Society of the Pacific, 1998, visited at 2003, http://www.eso.org/gen-fac/libraries/lisa3/ruschfejad.html

(٢) ibid.

إنجلترا وأستراليا وكندا واليابان وألمانيا وغيرها من الدول. وقد استخدمت الهيئات غير الرسمية القائمة على تطوير وتحديث بيئة الويب المايير المختلفة الداعمة لحقول مثل دبلن كور والتي تم تطويرها في اللغة المعياريا الجديدة XML. ويقترح الباحث استخدام هذه التحديثات في حقول ميتا لدعمملفات الصور (المستوى الثاني للاستخدام) في بناء تسجيلة معيارية للصور الرقمة، كما جاء في معيار دبلن كور على النحو التالي :- [١]

الجدول رقم (٢-٣) حقول الميتاداتا لوصف ملفات الصور لرقمية

DC. Title	Title of the Resource
DC. Creator	Author, Creator
DC. Subject	Subject, Keyword
DC. Descriptio	Annotation, Abstract, etc.
DC.Publisher	Publisher (Person or Institution)
DC. Contributc	Contributing Person or Institution
DC. Date	Date (see separate list of Sub-Elements "DC. Date")
DC. Type	Resource Type (according to a list of accepted terms)
DC. Format	Format, File Type, also Physical Medium
DC. Identifier	Resource Identification: URL, URN, ISBN, etc.
DC. Source	Resource (physical, digital) from which the current resource was derived, digitized, etc.
DC. Language	Language of the Resource
DC. Relation	Relationship to other Works
DC. Coverage	Geographic or Temporal Coverage
DC. Rights Cpy-right	Rights Management Statement (or Link to),

و يمكن أن تكون المعايير المستخدمة الملائمة لبيئة اللغة لعربية من هذا المعيار هي:-

(١) ibid.

الجدول رقم (٢-٤)

حقول الميتاداتا لوصف ملفات الصور الرقمية للمصادر العربية

DC. العنوان	عنوان المصدر
DC. المنشئ	الفنان، المصور (المؤلف)
DC. الموضوع	الموضوع أو الكلمة المفتاحية
DC. الوصف	المحتوى و المستخلص للموضوع
DC. الناشر	الناشر" الفرد أو المؤسسة"
DC. المشارك	الهيئة أو الفرد المشارك
DC. التاريخ	التاريخ" و هناك قائمة منفصلة لعناصر متعددة في التاريخ"
DC. النوع	نوع المصدر" طبقا للقائمة المحددة"
DC. الشكل	الشكل، نوع الملف الرقمي،الوسيط المادي
DC. محدد المصدر	محدد المصدر " و قد يكون مسار الملف، الرقم الدولي الموحد..."
DC. المصدر	المصدر "رقمي، تقليدي" الذي اشتق منه الملف الرقمي الحالي
DC. اللغة	لغة المصدر
DC. العلاقات	العلاقات البيليوجرافية مع الأعمال الأخرى
DC. التغطية	التغطية المكانية أو الزمنية للمصدر
DC. حقوق النشر	الحقوق المنظمة لنشر وتحميل المصدر

وحري بالذكر هنا أن المعيار المقترح السابق الذكر لإنتاج وإخراج ملفات الصور الرقمية على الويب يعمل على تغيير نمط التعامل بين محركات البحث وملفات الصور من الشكل غير المقنن إلى الشكل المقنن؛ مما ينتج عنه اختلاف في طبيعة المعالجة الفنية لملفات الصور من حقول تابعة لصفحات الويب غير منظمة إلى حقول محددة يمكن معها بناء معيار ثابت لتحرك برنامج الزاحف والمكشف في التعامل مع ملفات الصور على الويب. ويمكن متابعة ذلك كما سيلي في معالجة وبحث الصور الرقمية داخل محركات البحث في الفصول القادمة.

٢/٤/٢/٦ معيار RGL

وضع هذا المعيار مجموعة المكتبات البحثية Research Group Libraries (RGL)

عام ١٩٩٨، وعرف باسم RGL Preservation Metadata Elements. وتم بناء هذا المعيار اعتمادًا على العناصر الأساسية لمعيار Dublin Core.[١]

والجدير بالذكر أن كل هذه المعايير مازالت في مرحلة التجربة، وتتكون هذه المعايير من ثلاثة أجزاء رئيسة هي معلومات عن ملف الصورة ومحتوى الصورة والكيفية التي خرجت على أثرها الصورة. ويرى الباحث أن معياري الميتاداتا - دبلن كور هما أفضل المعايير في وصف الملفات الرقمية للصور، وسوف يتم دراسة عناصره مشروحة في وصف عناصر الصورة المادية والموضوعية في الفصل الثالث.

٢/٤/٣ عناصر محورية في وصف الصور الرقمية

أكد Michael Day أهمية وصف جوانب محددة في صياغة معايير وصف ملفات الصور الرقمية، وتتعلق أغلب هذه الجوانب بإدارة الحاسب الآلي لملفات الصور إلى جانب ما يخص مستخدم هذه الملفات. ومنها:-[٢]

٢/٤/٣/١ المعلومات الفنية التي تتعلق بعرض الصور الرقمية : حيث تذكر نوع ملف الصورة، مدى عمق الصورة وكثافتها إلى جانب الشكل التقني للصورة سواء كان (TIFF, GIF.).. أو غيرها.

٢/٤/٣/٢ المعلومات حول بناء الصورة الرقمي : حيث تذكر جوانب حجم ملف الصورة، ومجهز الصورة provider، ونوع الماسح الضوئي scanner والأبعاد داخل الصورة.

٢/٤/٣/٣ المعلومات حول دقة وجودة الصورة الرقمية : حيث يذكر المسئول عن رقمنة الصورة digitization، ومثال ذلك (صور رقمية تحت رعاية المتحف المصري) أو بواسطة شخص غير تابع لهيئة.

٢/٤/٣/٤ المعلومات عن الصورة الأصلية : حيث تذكر المعلومات عن مصدر الصورة وطبيعته، ويمكن أن تحتوي هذه العناصر على وصف سابق أو تصنيف موضوعي للصورة الأصلية.

---

(١) Metadata: Image Description, Indexing and Cataloguing. op.ct.

(٢) Day,Michael. Metadata for images: emerging practice and standards, UKOLN: The UK Office for Library and Information Networking, 1999, visited at 12/2/2004, http://www.ukoln.ac.uk/metadata/presentations/cir99/paper.html

٢/٤/٣/٥ المعلومات حول توثيق الصور : حيث يسترجع مستخدمو الصور المئات من ملفات الصور بسهولة ويسر؛ إلا أنهم في الوقت ذاته غير واثقين من خلو الصورة من التغيرات أو التعديلات غير المسئولة. ويمكن التغلب على هذه المشكلة من خلال استخدام تقنيات حق النشر مثل العلامات المائية داخل ملفات الصور.

٢/٤/٣/٦ المعلومات حول حقوق النشر : تفيد هذه المعلومات في التعامل مع ملفات الصور أكثر من مرة، وتحفظ هذه المعلومات حقوق الاستخدام سواء للناشر أو لمنتج الصورة الرقمية.

ويقترح Howard Besser أن تكون جوانب الوصف السابقة محورية في إنتاج معايير المعالجة الفنية للصور الرقمية في مختلف أنواعها، و أشكال ملفاتها.

٢/٥ الخلاصة

تختلف تطبيقات الصور الرقمية بحسب كل من مجال الاستخدام وتقنيات أو أماكن الاستخدام، كما وضح ذلك في استخدامات الأغراض البحثية و العلمية والتاريخية الثقافية إلى الفضائية وصور وكالة ناسا. أيضا هناك استخدامات الصور في مجالات التسويق و العروض الفنية إلى جانب تصميم صفحات الويب. وتقع الصور الرقمية الطبيعية كورقة من فرع من شجرة وسائط حمل المعلومات المرئية بصفة عامة. وبالإضافة إلى ذلك فإن أشكال الصور الرقمية ليست موحدة، وإنما تتباين بحسب التقنيات المستخدمة في إنشاء ملفات الصور الرقمية على الويب. و قد عمل هذا الفصل أيضا على استعراض المعايير المستخدمة للمعالجة الفنية لصور الويب، مع التمهيد للتركيز على أفضلها في الفصل التالي من الدراسة.



الفصل الثالث

معالجـــة الصـــور

٣/٠ التمهيد

٣/١ الوصف المادي والموضوعي للصور الرقمية

٣/٢ معالجة اللغة العربية في نظام استرجاع الصور

٣/٣ الخلاصة



معالجة الصــــور

٣/٠ التمهيد

تحتاج الصور الرقمية في بيئة الويب إلى مجموعة من عناصر الوصف الفني المادي والموضوعي التي تلائم التعامل مع ملفات الصور الرقمية. وتؤثر عملية الوصف الفني للصور الرقمية في أساليب التنظيم والمعالجة التي يتوقف عليها استدعاء وبحث الصور الرقمية داخل محركات بحث الويب و المحركات المحددة. و لكن إلى أي مدى قد تعاملت الدراسات مع المعالجة الفنية للصور الرقمية، وكيف يمكن الخروج بالتسجيلة الببليوجرافية الملائمة لملفات الصور الرقمية على الويب.

٣/١ الوصف المادي والموضوعي للصور الرقمية

٣/١/١ الوصف المادي للصور الفوتوغرافية الرقمية

إن الحديث عن المكونات المادية للصورة جاء نتيجة الحاجة لوصف هذه الصورة وصفًا ببليوجرافيًا، ذلك إنما ينقسم هذا الوصف الببليوجرافي إلى الشقين الأساسيين وهما المادي والموضوعي. والأمر يختلف بين وصف مادة الصورة في البيئة التقليدية وعنها في البيئة الرقمية حيث تفرض كل منها عناصر تلائم خصائص الاسترجاع فيها. فمثلا تفرض البيئة الرقمية الحديث عن نوع الملف GIF, JPEG, PNG... إلخ، كما أن التسجيلة في البيئة التقليدية تختلف عن التسجيلة في البيئة الرقمية، فالأولى لا يتم الاسترجاع فيها إلا بالمدخل الذي رتبت على أثره بطاقة الوصف الببليوجرافي الذي يكون في أغلب الأحيان مدخل المؤلف أو العنوان. أما الأخيرة فيمكن أن تسترجع بكل عناصر التسجيلة.

واعتمادًا على أن الصورة يتم تخزينها وعرضها في البيئة الرقمية، فإن عناصر وصف تسجيلات الصور تأتي من كل عناصر ملفات الصور مثل حجم الملف، وألوان الصورة، وشكل الصورة داخل الملف. وقد أوضح Jack Christine أن هذه العناصر مع غيرها مثل المصور، والزمن الذي التقطت فيه الصورة، ومالك الصورة، والبنية الرقمية التي وضعت بها الصورة، كل ذلك يمكن أن يشكل التسجيلة الكاملة لوصف وسيط الصورة الفوتوغرافية.(١) وعلى ذلك فيمكن البحث عن لوحة بألوان الزيت رسمها ليوناردو دافنشي إذا ما تم تحديد تاريخ الصورة ونوع الألوان ومكان الوجود بالإضافة إلى موضوع هذه اللوحة والتي يشكل الموضوع فيها أهم عناصر استرجاعها.

وتسمى العناصر السابقة التي تصف الصورة كما أشرنا في الفصل الثاني باسم الميتاداتا، وتعمل هذه العناصر عمل النص المساعد في استرجاع الصورة المكونة من حقول وكلمات مفتاحية ومسار موقع الملف في حالة صور الويب، بالإضافة إلى امتداد الملف الذي يوجد في نهاية مسار ملفات الصور، هذا ما تستند إليه محركات البحث في حالة البحث عن ملفات الصور. وتختلف طبيعة المعايير في التسجيلات الرقمية عنها في التسجيلات التقليدية حيث لا يمتد الأمر في التسجيلات الرقمية إلى الأبعاد بين الحقول أو علامات الفصل بين الحقول وما إلى ذلك من روابط بين الحقول. فالأمر لا يتعدى اختيار الحقول التي ستوضع في التسجيلة والعناصر التي تمثل هذا الوسيط، وهذا هو ما توجد عليه النظم الآلية الحالية في مجال المكتبات .

ولقد اقترح Jane Hunter نظامًا لاسترجاع الصور يعتمد على تكشيف الصور ووصفها بالكلمات المفتاحية بالإضافة إلى الاعتماد على شكل معين من ملفات الصور الرقمية هو PNG،(٢) وهذا النظام قد تم صياغته وتحديد معايير مصطلحاته عن طريق مجموعة العمل لمنظمة Dublin Core Metadata Initiative (DCMI)(٣). وقد

---

(١) Cristine, Jake. stat of the arts: current applications for indexing images, the author, 1999, visited at 23/9/2002, http://portal.acm.org/citation.cfm?id=614788

(٢) Hunter, Jane, op.cit., p.10

(٣) هذه إحدى المنظمات في المملكة المتحدة والتي تعمل في مجال استرجاع مصادر المعلومات على الويب، ذلك من خلال صياغة النظم المثلى في استرجاع المصادر بالإضافة إلى صياغة المعايير اللازمة لتحقيق افضل أداء لاسترجاع المعلومات. ويمكن الرجوع إلى موقع المنظمة http://uk.dublincore.org/

وضع Hunter في هذا الوصف حقولا معيارية للوصف الببليوجرافي للصور. ويرى الباحث أن هذه الحقول قد جمعت كل ما يمكن أن يوجد في الصور من سمات مادية يمكن استرجاع الصورة الفوتوغرافية من خلالها. والجدير بالذكر هنا أن هذه التسجيلة لا تلائم سوى الصور الفوتوغرافية وليس كل أشكال المادة المصورة التي تندرج تحت مصطلح (Image) وهذه الحقول كما يلي:

الجدول رقم (٣-١) التسجيلة الإنجليزية للصور الرقمية في المستوى الأول

Top Level Metadata Description for Complete Image

Title	A selector and his family, probably in the Beenleigh district, 1872
Creator	William Boag
Subject	Photograph collection - Queensland
Description	The difficulties faced by a family in the Queensland bush included poor roads, an unreliable mail service and dense, vine-matted scrub. For many years, a selector's staple diet was salted meat (salt horse) and pumpkins. For several months, a woman and her children might be alone in their stringy-bark hut while her husband went off to split shingles or to earn extra money on a cattle property.
Date.created	1872
Date.recordCreated	1996
Date.placedOnline	1997
Publisher	State Library of Queensland
Type	image.photograph
Format	image/jpg
Format.fileSize	50.6Kb
Format.dimensions	672 x 512
Format.colorpalette	Grayscale
Identifier	http: //archive.dstc.edu.au/RDU/SLQ/boag/20248.jpg
Source	BOAG negative no. 906
Language	En
Relation.isPartOf	http://www.slq.qld.gov.au/jol/boag.htm
Relation.hasParts	Region1
Relation.hasFormat	http://archive.dstc.edu.au/RDU/SLQ/boag/20248.gif
Coverage	Beenleigh region, Queensland, 1872
Rights	http://www.slq.qld.gov.au/cright.htm

الجدول رقم (٣-٢) التسجيلة الإنجليزية للصور الرقمية في المستوى الثاني

Secondary Level Metadata for Region1

Identifier	Region1
Title	Annie Dickson
Description	Wife of James Dickson and mother to their 13 children.
Coverage.rect	495,207,546,263
Relation.isPartOf	http: //archive.dstc.edu.au/RDU/SLQ/boag/20248.jpg

ويرى الباحث أن شكل التسجيلة السابقة يمكن أن يستخدم في البيئة العربية بالشكل التالي :-

الجدول رقم (٣-٣) التسجيلة العربية للصور الرقمية في المستوى الأول

العنوان	رجل وعائلته في مقاطعة Boonleigh في عام ١٨٧٢
المصور	William Boag
الموضوع	مجموعات صور-كوينزلاند
الوصف	بعض من الصعوبات التي تواجه العائلة و منها الطرق غير الجيدة بالإضافة إلى الخدمات السيئة للبريد و المعيشة، وعلى الزوجة و أولادها انتظار الأب لإحضار بعض المال.
تاريخ الإنشاء	١٨٧٢
تاريخ إنشاء التسجيلة	١٩٩٦
تاريخ التحميل على الخط المباشر	١٩٩٧
الناشر	مكتبة ولاية كوينزلاند
النوع	صورة فوتوغرافية
الشكل	صورة من نوع jpg
حجم ملف الصورة	50.6Kb
أبعاد الصورة	672 x 512
اللون العام للصورة	اللون الرمادي

تابع ـ الجدول رقم (٣-٣) التسجيلة العربية للصور الرقمية في المستوى الأول

محدد المصدر الموحد	http: //archive.dstc.edu.au/RDU/SLQ/boag/20248.jpg
المصدر	الفيلم السالب رقم ٩٠٦
اللغة	الإنجليزية
العلاقة الببليوجرافية	http://www.slq.qld.gov.au/jol/boag.htm جزء من صفحة
نوع العلاقة	Region1علاقة أولية
مسار العلاقة من الصفحة	http://archive.dstc.edu.au/RDU/SLQ/boag/20248.gif
التغطية	, مقاطعة Beenleigh ولاية Queensland, 1872
حقوق النشر	http://www.slq.qld.gov.au/cright.htm

الجدول رقم (٣-٤) المستوى الثاني وصف اللغة العربية

المحدد	Region1
العنوان	Annie Dickson
الوصف	زوجة لـ James Dickson و أم لـ١٢ طفلًا
رقم التغطية في المكتبة	495,207,546,263
العلاقة الببليوجرافية	http: //archive.dstc.edu.au/RDU/SLQ/boag/20248.jpg

١/١/١/٣ خصائص التسجيلة في معيار الميتاداتا

تمثل عناصر التسجيلة السابقة مواصفات واحدة من مجموعة صور، وما يمكن الحديث عنه هنا هو حقول هذه الصورة التي تشكل كل عناصر وصف الصورة وملفات صور الويب في المستوى الكامل للوصف الببليوجرافي، ويمكن هنا تناول العناصر السابقة بالشرح فيما يخص آليات استرجاع الصور وما تسهم به هذه العناصر في وصف الصورة واسترجاعها في مستواها الأول على النحو التالي:-

أولا : أن كل العناصر السابقة تصف الشكل المادي للصورة وللملف داخل قاعدة البيانات؛ إلا أن التسجيلة قد شملت ضمن ما شملت من عناصر حقل "Description" ويعنى هذا الحقل بوصف المحتوى الموضوعي للصورة. وقد

لاحظ الباحث أن الوصف هنا لم يكن في كلمات مفتاحية منفردة وإنما جاء الوصف في شكل ملخص Summary يحتوي على وصف الموضوع والعناصر المادية والموضوعية داخل الصورة، وهذا يتوافق مع ما سبق ذكره من صعوبة تغيير لغة الصورة إلى لغة نص وتحويل الرؤية إلى كلمات تكون هي الأكثر دلالة على الموضوع. كما أن هذا أيضا يتوافق مع طبيعة العنصر البشري في وصف محتوى المادة المصورة.

ثانيًا : لقد تم وضع حقل الموضوع Subject؛ إلا أن رأس الموضوع لم يكن لما هو بداخل الصورة وإنما لتحديد الصورة في إطارها العام حيث يذكر "Subject: Photograph collection - Queensland" ويرى الباحث أن هذا الحقل لما هو بداخل الصورة وليس لما تقع الصورة بداخله. حيث يأتي رأس الموضوع المعبر عن الصورة (الحياة الأسرية) أو الأبناء ـ رعاية... الخ.

ثالثًا : لقد ذكر معيار الوصف هنا ثلاثة أنواع من تاريخ الوسيط هي تاريخ الإنتاج، وتاريخ الوصف، وتاريخ الإتاحة على الخط المباشر؛ إلا أن أهم هذه التواريخ هي تاريخ الإنتاج ذلك لأن بعض الباحثين قد يهمهم بالدرجة الأولى الملامح التاريخية لهذه الفترة المتمثلة في محتويات الصورة.

رابعًا : إن الحديث عن الوصف المادي لا يكون هنا عن الصورة وإنما الحديث عن الملف الإلكتروني الذي يحتوي الصورة، وذلك في إعطاء نوع الملف وحجم الملف ومكان وجود هذا الملف على الخط المباشر أو في الويب. وكل هذه البيانات وإن كانت تفيد الباحث في الوصول إلى الصور التي يريدها؛ إلا أنها تهم بالدرجة الأولى محركات بحث الصور؛ ذلك لأن محرك البحث يبحث داخل كل ملفات الويب عن هذا النوع من الملفات قبل أن تكون متاحة للباحثين، وعلى ذلك فإن النظر هنا يكون إلى برامج الزحف الآلية Spidering التي تقوم باقتناء وإضافة ملفات الصور إلى قواعد بيانات محركات البحث.

خامسًا : يأتي الجزء الخاص بحماية الملكية الفكرية والنشر عن طريق ذكر موقع أو صفحة المجموعة المصورة التي تنتمي إليها هذه الصورة. وجاء وجود هذا الحقل في تسجيلة الصورة نتيجة للكتابات الكثيرة التي تناولت حق

الملكية للصور، ويعد هذا أيضًا نابعًا من الوجهة التجارية التي نشأت بها قواعد بيانات الصور ومحركات الصور أيضًا.

سادسًا : تساعد عناصر التغطية اللغوية والجغرافية بذكر المنطقة التي تضمها الصورة واللغة التي وصفت بها الصورة في تعدد استخدامات الصورة، وليس فقط الاستخدام العام لها أو الفني.

أما المستوى الثاني من الوصف الببليوجرافي للصور فيقترح الباحث أن يصاحب الصورة عند عرض جميع النتائج التي تم استرجاعها من كل نظام الاسترجاع حيث يمكن أن يكون له التأثير في ترجيح مجموعة من الصور عن الأخرى، وذلك بخلاف المستوى الأول للوصف الذي هو بمثابة التسجيلة الكاملة لتحليل الصورة ماديًا وموضوعيًا.

و قد وضع Garry Forger وصفًا لعناصر تسجيلة Dublin Core تتضمن مصادر بيانات الحقول وعناصرها كما يلي(١) :-

الجدول رقم (٣-٥)

التسجيلة الكاملة الأجنبية للصورة الرقمية العامة في بيئة الويب

ELEMENT	DESCRIPTION
TITLE	The name given to the resource by the CREATOR or PUBLISHER.
CREATOR	The person(s) or organization(s) primarily responsible for the intellectual content of the resource. For example, authors in the case of written documents, artists, photographers, or illustrators in the case of visual resources.
SUBJECT	The topic of the resource, or keywords or phrases that describe the subject or content of the resource. The intent of the specification of this element is to promote the use of controlled vocabularies and keywords. This element might well include scheme-qualified classification data (for example, Library of Congress Classification Numbers or Dewey Decimal numbers) or scheme-qualified controlled vocabularies (such as MEdical Subject Headings or Art and Architecture Thesaurus descriptors) as well.
DESCRIPTION	A textual description of the content of the resource, including abstracts in the case of document-like objects or content descriptions in the case of visual resources.
PUBLISHER	The entity responsible for making the resource available in its present form, such as a publisher, a university department, or a corporate entity. The intent of specifying this field is to identify the entity that provides access to the resource.

(١) Forger, Garry. Image, University of Arizona Library, 2002?, visited at 5/10/2003, http://dizzy.library.arizona.edu/sitesrch/train.html

تابع - الجدول رقم (٣-٥)

التسجيلة الكاملة الأجنبية للصورة الرقمية العامة في بيئة الويب

ELEMENT	DESCRIPTION
CONTRIBUTOR	Person(s) or organization(s) in addition to those specified in the CREATOR element who have made significant intellectual contributions to the resource but whose contribution is secondary to the individuals or entities specifed in the CREATOR element (for example, editors, transcribers, illustrators, and convenors).
DATE	The date the resource was made available in its present form. The recommended best practice is an 8 digit number in the form YYYYMMDD as defined by ANSI X3.30-1985. In this scheme, the date element for the day this is written would be 19961203, or Dec ember 3, 1996. Many other schema are possible, but if used, they should be identified in an unambiguous manner.
TYPE	The category of the resource, such as home page, novel, poem, working paper, technical report, essay, dictionary. It is expected that RESOURCE TYPE will be chosen from an enumerated list of types.
FORMAT	The data representation of the resource, such as JPEG or TIFF image. The intent of specifying this element is to provide information necessary to allow people or machines to make decisions about th e usability of the encoded data (what hardware and software might be required to display or execute it, for example). As with RESOURCE TYPE, FORMAT will be assigned from enumerated list. In principal, formats can include physical media such as books, serials, or other non-electronic media.
IDENTIFIER	String or number used to uniquely identify the resource. Examples for networked resources include cs and URNs (when implemented). Other globally-unique identifiers, such as International Standard Book Numbers (ISBN) or other formal names would also be candidates for this element.
SOURCE	The work, either print or electronic, from which this resource is derived, if applicable. For example, an html encoding of a Shakespearean sonnet might identify the paper version of the sonnet from which the electronic version was transcribed.
LANGUAGE	Language(s) of the intellectual content of the resource. Where practical, the content of this field should coincide with the NISO Z39.53 three character codes for written languages.
RELATION	Relationship to other resources. The intent of specifying this element is to provide a means to express relationships among resources that have formal relationships to others, but exist as discrete resources themselves. For example, images in a docume nt, chapters in a book, or items in a collection. A formal specification of RELATION is currently under development. Users and developers should understand that use of this element should be currently considered experimental.
COVERAGE	The spatial locations and temporal duration characteristic of the resource. Formal specification of COVERAGE is currently under development. Users and developers should understand that use of this element should be currently considered experimental.
RIGHTS	The content of this element is intended to be a link (a URL or other suitable URI as appropriate) to a copyright notice, a rights-management statement, or perhaps a server that would provide such information in a dynamic way. The intent of specifying this field is to allow providers a means to associate terms and conditions or copyright statements with a resource or collection of resources. No assumptions should be made by users if such a field is empty or not present.

الجدول رقم (٣-٦)

التسجيلة الكاملة العربية للصورة الرقمية العامة في اللغة العربية

العنصر	الوصف	مصدر العنصر	شكل التسجيل	مثال
العنوان	هو اسم الوسيط الذي وضع من جانب المصور أو الناشر	يؤخذ من الأصل		حديقة الأزهار
المؤلف "المصور"	الشخص المسئول الأول عن محتوى الصورة : المصور	يؤخذ من الأصل	يبدأ باسم العائلة في الأسماء الأجنبية والاسم الأول في الأسماء العربية	كلارك فريد
الموضوع	مصطلح أو عبارة لوصف محتوى الصورة	مختلف المصادر	بشكل المكانز أو رؤوس الموضوعات	الحرب العالمية الثانية العرض العسكري
مستخلص	وصف نصي موجز لمحتوى الصورة يعرض الموضوعات الأساسية بها	من الوسيط نفسه أو بواسطة مكشف	جمل واصفة	تدور حول المقاومة الفلسطينية لقوات الاحتلال
الناشر	الفرد أو الجهة المسئولة عن إظهار الصورة في الشكل الذي خرجت عليه	من الوسيط نفسه		الأرشيف القومي المصري
الموزع	الفرد أو الجهة التي تعمل على إيصال المحتوى إلى المستخدمين	من الوسيط نفسه	اسم الفرد أو الهيئة الشائع	

تابع الجدول رقم (٣-٦)

التسجيلة الكاملة العربية للصورة الرقمية العامة في اللغة العربية

العنصر	الوصف	مصدر العنصر	شكل التسجيل	مثال
التاريخ	تاريخ وضع الوسيط في الشكل الحالي	المكتبة الوطنية ١٩٧٨	اليوم/الشهر/العام	١٩٧٨/١١/٢٤
النوع	نوع الوسيط سواء صورة فوتوغرافية، شريحة، لوحة...	الوسيط		صورة فوتوغرافية، ملونة أو أبيض وأسود
الشكل	الشكل الرقمي لملف الصورة	من خصائص الملف الرقمي وامتداد عنوانه		GIF, JPG
المحدد "رقم الطلب"	هو كلمة أو رمز لتحديد الصورة داخل القاعدة		شكل الأرقام أو المختلط والهجائي	أ/١٥/٩٦
المصدر الأصلي	الشكل الأصلي للصورة. لوحة أو مطبوع			
اللغة	لغة محتوى الصورة	الوسيط أو الملف	ثلاثة حروف أو اسم اللغة	
المنطقة	منطقة محتوى الصورة			شبه الجزيرة العربية
الحجم	مساحة ملف الصورة بالكيلو بايت			100 k.b.
حق النشر	حقوق النشر المتعلقة باستخدام الصورة ونسخها وتحميلها			المكتبة القومية المصرية
البريد الإلكتروني	البريد الإلكتروني لمصور أو مالك الصورة			oooo@yahoo.com
مسار ملف الصورة	موقع أو محدد مسار ملف الصورة على الويب		www.ifla.org/conferences/leader.gif	

العنوان:

الفنان، المصور (المؤلف):

الموضوع:

وصف الموضوع (المحتوى):

تاريخ التقاط الصورة:

تاريخ صنع الصورة الرقمية:

تاريخ وضع التسجيلة:

الناشر:

شكل ملف الصورة:

حجم ملف الصورة:

أبعاد الصورة:

اللون:

تعريف الصورة: مسار الملف:

مصدر الصورة على الويب : في حالة محرك البحث: موضع التقاط الصورة من الويب.

لغة ملف الصورة : في حالة المحركات متعددة اللغات:

تغطية الصورة: المكان:

حقوق نشر الصورة:

وتمثل الحقول السابقة التسجيلة المقترحة لصياغة ملف الصور الرقمية، إلى جانب الأخذ بها في تكشيف محركات البحث لملفات الصور، من خلال تعيين المصطلحات من حقول تسجيلة ملف الصورة إلى التسجيلة داخل نظام الاسترجاع لمحرك البحث.

٢/١/٣ التحليل الموضوعي للصور

تناول معيار الوصف السابق تحليل الشكل المادي للصورة. أما تحليل المحتوى، فانه يأخذ جانبًا آخر، حيث لا يعتمد الأمر هنا على وصف عناصر

وأشكال مادية يتم وصفها كما هي وإنما يعتمد على عناصر موضوعية ومعانٍ يمكن أن تفهم بأكثر من زاوية لرؤية المعنى. ويسمى استرجاع الصور في الحالة التي لا يستخدم فيها النص باعتباره عامل استرجاع للصور باسم استرجاع الصور المبني على المحتوى Content Based Images Retrieval (CBIR)، وقد ظهر هذا المفهوم على يد Kato عام ١٩٩٢م عندما بدأ العمل في اتجاه سترجاع الصور للصورة ويكون برنامج نظام الاسترجاع ذا خصائص تمنحه الندرة على تحليل مكونات الصورة الأساسية وهي اللون والشكل shape والبنية texture.[١] ولا تدخل للنص هنا في الاسترجاع سواء بالوصف أو بالكلمات المفتاحية، إنما يعتمد فقط على الخصائص والأساليب الآلية الخالصة. وهذا النوع من لاسترجاع تكون الحاجة له حيث يكون فيه استخدام الواصفات أو النص غير الوارد مثل التعرف إلى بصمات شخص ما ومحاولة مطابقتها على بصمات أخرى في قاعدة بيانات. أو التعرف إلى شكل فيروس أو بكتيريا غير معروفة ومطلوب مطابقتها على مواد أخرى شبيهة في قاعدة بيانات .

أما استرجاع الصور المبني على النص المصاحب للصورة حيث تكون الصورة فقط جزءًا من التسجيلة فيسمى Text Based Images Retrieval (TBIR). وهذا هو الأنسب في استخدامه على الويب لكن هذا الجزء يساهم في الاسترجاع فقط بصفاته الشكلية التي تم تحويلها إلى كلمات ونص مصاحب سواء كانت كلمات واصفة أو عبارات واصفة أو جملًا شارحه. وهذا النوع من البرامج هو ما تحاول هذه الدراسة العمل على إيجاد مواصفات تزيد من كفاية عملية البحث والاسترجاع داخله .

وينقسم استرجاع الصور المبني على المحتوى والموضوع إلى نوعين من الاسترجاع. أولهما: تقسيم كل مجموعة صور في مكتبة الصور إلى موضوعات رئيسة ثم إلى موضوعات فرعية ثم إلى فروع الفروع وهكذا، أما الشكل الآخر فهو استرجاع مبني على البحث بالكلمات المفتاحية والذي من الممكن أن يكون في شكل واصفات منفردة أو عبارات غير متكاملة أو جمل في شكل ملخص

---

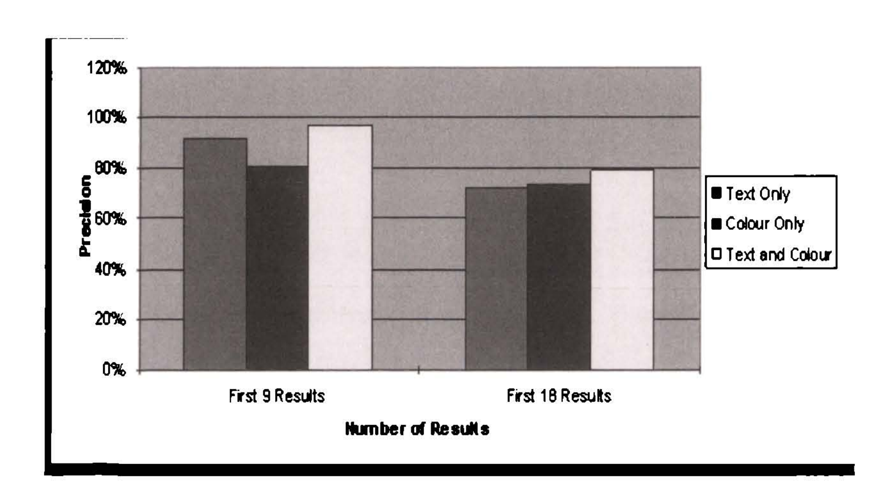
(١) Mattison, david. Images of History on the Web. informat:ion today,inc, 2002, visited at: 15/10/2003, http://www.infotoday.com/searcher/may02/mattison.htm

شارح لموضوع ومحتوى الصورة. وهذان النوعان هما اللذان يستخدمان في محركات البحث سواء التقسيم الموضوعي كما في أدلة الصور أو البحث الحر كما في محركات البحث.

وقد اتجه نظر باحثي استرجاع الصور إلى زاوية أخرى مؤداها تحقيق التكامل بين نوعي استرجاع الصور، يأتي ذلك فيما اقترحته دراسة Guojun Lu and Ben Williams في تصميم نظام متكامل لاسترجاع الصور يجمع بين استرجاع الصور المبني على النص واسترجاع الصور المبني على العناصر الشكلية[(١)]. ونمت فكرة الدراسة نتيجة لما يواجهه كلا النوعين من قصور في تحقيق الكفاية المثلى ؛ حيث أشار الباحثان إلى أن وصف استرجاع الصورة باستخدام النص المصاحب قد لا يصل إلى الحد الذي يعبر فيه المكشف عن كل معاني الصورة وموضوعاتها، بما ينتج عنه استرجاع غير دقيق للصور، وعلى الجانب الآخر، فإن الاسترجاع باستخدام محتوى الصور والعناصر الشكلية والمادية للصورة ينتج عنه قصور في تحديد الموضوعات والمعاني التي تحويها الصورة. ويكون الاسترجاع هنا على نحو طرح الأسئلة البحثية في شكل أسئلة نصية ثم تلقي النتائج واختيار الوحدات المطابقة للموضوع منها، وبعد ذلك يأتي دور الاعتماد على محتوى الصورة في مطابقة كل الوحدات الأخرى مع الوحدات المختارة في النتائج الأولية. والاعتماد على استرجاع الصور في المرحلة الثانية لا يكون إلا على اللون فقط وليس كل خصائص المحتوى الثلاثة (اللون والشكل والبنية) . وقد علل الباحثان ذلك بأن الاعتماد على الشكل والبنية يحتاج إلى معالجة ذات درجة عالية في استرجاع الصور، إضافة إلى أن اللون هو أكثر الخصائص المؤثرة في استرجاع الصور. وقد أجرى Lu و Williams دراسة تجريبية لاستخدام نظام متكامل لاسترجاع الصور يعتمد على كل من النص واللون، وأثبتت التجارب على الاسترجاع باللون بمفرده، والنص بمفرده، والنص واللون معًا أن تكامل النص واللون قد حقق أعلى درجة تحقيق بالقياس على الـ ٩ وحدات الأولى يرمز لها بالرمز (ن) والقياس على الـ ١٨ وحدة الأولى ؛ وكانت النتائج كما يلي[(٢)]:-

(١) Lu, Guojun and Williams, Ben. An Integrated WWW Image Retrieval System, the author, 1999, visited at 6/10/2002, http://www.onlinemag.net/OL2000/net9.html

(٢) ibid.



الشكل رقم (٣-١)

يوضح نتائج التحقيق للون بمفرده والنص بمفرده واللون والنص معًا

ولقد لاحظ الباحث من دراسة Lu و Williams تأكيدهما على الاستعانة بالمعاني والموضوعات الداخلية للصورة في استرجاعها، في الوقت الذي تركزت فيه أفكار الكتابات الأخرى مثل K.W. Tobin, T.P. Karnowski على الاستعانة بالشكل واللون وبنية الصورة في بناء نظم استرجاع الصور الرقمية(١).

واعتمادًا على هذا الهدف، فإن ما يلي يدور حول المعالجة الفنية المثلى للصور من خلال النص المصاحب لها، متخذة من بيانات ملفات اللغة المعيارية HTML أساسًا لتكوين وحدات معالجة الصور. ولأن هذا النوع من الملفات تختلف فيه أهمية الكلمات بحسب أماكن تواجدها داخل الملف، فقد اهتمت الدراسة بكيفية صياغة وصف الصور في كلمات محددة والأماكن التي تأخذها داخل ملف الصورة من نوع HTML. وعلى ذلك فإنه يؤخذ في الاعتبار مكان الكلمات المفتاحية الذي يعبر عن وزنها في استرجاع الصورة. وفيما يلي يتم تحديد أنواع المصطلحات ووزنها باعتبارهما المتحكمين في استرجاع صورة دون أخرى داخل ملف اللغة المعيارية.

---

(١) Karnowski, T.P. image retieval, Image Science and Machine Vision, 2002, visited at 12/5/2004, http://www-ismv.ic.ornl.gov/projects/video.htm

أولا : مجموعة كلمات الرموز المعيارية Meta data وهي الكلمات المرتبطة بتعريف صفحات الويب من لدن مصمميها، وتأخذ وزن (٠,٢) في تأثيرها على استرجاع الصور.

ثانيًا : مجموعة كلمات العنوان title وهي الكلمة أو الكلمات القليلة المعبرة عن موضوع الصورة، وتأخذ وزن (٠,٢) في تأثيرها على استرجاع الصور.

ثالثًا : مجموعة كلمات وهي الكلمات الأساسية الأولى المصاحبة للصورة إن وجدت Heading، وتأخذ وزن (٠,٢) في تأثيرها على استرجاع الصور.

رابعًا : مجموعة كلمات Same Paragraph وهي مجموعة الكلمات التي تم جمعها خصيصًا للتعبير عن موضوع هذه الصورة بذاتها، وتأخذ وزن (٠,٣) في تأثيرها على استرجاع الصور.

خامسًا : مجموعة كلمات مسار ملف الصورة وهي أكثر الكلمات تعبيرًا عن المحتوى والموضوع كالسيارة والمواصلات والعميد. ويأخذ وزن (٠,٤) تأثيره. إلا أن هذا قد لا يناسب اللغة العربية باعتبار أن المسار يكتب باللغة الإنجليزية، ولكن في حالة ارتباط محرك البحث بقاموس يقوم بالترجمة الآلية إلى العربية فإنه يمكن الاستفادة من هذا المعيار.

سادسًا : **مجموعة كلمات النص المتردد** Alternate text وتعرف هذه الكلمات بأنها كلمات توضع في جانب خاص لها يسمى alt component، حيث تعطي أقرب وصف لمحتوى الصورة. وتأخذ وزن (٠,٤) في تأثيرها على استرجاع الصور.

سابعًا : مجموعة مصطلحات النص المعتمد Anchor text وهو النص الذي يدعم من قبل المصمم عند إنشاء الملف يوضع داخل محارف <a> and ، لكي تأخذ كلماته درجة أكبر في التأثير عن غيرها من كلمات ملف الصورة، ويأخذ (٠,٤) في تأثيره على استرجاع الصور.

ثامنًا : مجموعة المصطلحات الأخرى التي تدخل في تركيب ملفات الصور في بيئة HTM، وتأخذ وزن (٠,٢) في تأثيرها على استرجاع الصور.

وعلى النحو السابق فيرى الباحث أن مجموعات الكلمات لا يجب أن تستخدم فقط في استرجاع الصور بالكلمات، وإنما أيضا الوزن النسبي لهذه الأنواع

من الكلمات حسب أماكنها الذي يؤثر في ترتيب النتائج بالإضافة إلى إصباغ النتائج بدرجة أعلى من التحقيق. هذا بالإضافة إلى أن الدراسة قد انصبت على مناقشة المصطلحات والأماكن داخل ملف الصورة، ولم تتحدث عن مستويات التكشيف، ولا المصطلحات المستخدمة أو العبارات الواصفة. ومن ثم فإن أهمية وزن مصطلحات ملف الصورة يمكن أن تتعدى مرحلة البحث إلى التأثير في مرحلة الاقتناء والإضافة داخل محركات البحث، فإذا ما تم تدعيم برنامج الزاحف أو العنكبوت لالتقاط ملفات الصور حسب مجموعة كلمات محددة، مثل مجموعة كلمات العنوان، فإن ذلك يمكن أن يكون له أثر كبير في أهمية المجموعات التي سيقتنيها محرك البحث داخل قاعدة بياناته.

أما الأمر الآخر والذي يمكن استخلاصه من الدراسة السابقة، أنه من الممكن أن تستخدم العناصر السابقة في مرحلة أخرى خلاف مرحلة البحث والاسترجاع ؛ حيث يمكن الاستفادة منها في إنتاج ملفات الصور بحسب مواصفات محددة وتوزيع ثابت للكلمات المكونة للنص المصاحب للصورة، مما يمكن أن يؤدي بدوره إلى إدخال خصائص التكشيف الآلي لملفات الصور مع التأكد من جودة هذا التكشيف في اختيار كلماته المفتاحية.

كما يمكن أن تستخدم العناصر الأخرى فيما يسمى البحث في النتائج أو استخدام المرشحات لتنقية نتائج البحث. فمثلا يمكن استخدام البحث في مجموعات الكلمات كآلية بحث معتمدة في حالة البحث في نتائج البحث. كما يمكن استخدام البحث في المسار كآلية بحث في المرشحات.

وتنقسم نظم استرجاع الصور المعتمدة على تحليل المحتوى في شكل كلمات وهو النص المصاحب من حيث بحث ملفات الصور، إلى نوعين أساسين هما : البحث باستخدام التقسيم الموضوعي، والبحث باستخدام الكلمات المفتاحية، وهما على النحو التالي:-

١/٢/١/٣ البحث بالتقسيم الموضوعي

يأخذ البحث بالتقسيم الموضوعي أكثر من مسمى، حيث يطلق عليه أيضًا اسم التصفح. كما أن تطبيق هذا البحث يختلف من مجموعة إلى أخرى، فبينما تقوم بعض مكتبات الصور بالاعتماد على أحد أنظمة التكشيف أو التصنيف في عمل

تقسيم موضوعي من الأقسام العامة إلى الأقسام الأكثر تخصصيًا، تقوم بعض المكتبات الأخرى بعمل أنظمة خاصة بها. والعمل في مجال تنظيم الصور يسعى دائما إلى هدف التلاقي مع مستخدمي قواعد بيانات الصور في تصور محتوى الصورة الذي يعطي إمكانية التنظيم، وفي الوقت نفسه يعطي الباحث ما يريد من معلومات مرئية محددة. ويعتمد أسلوب التصفح على شكلين أساسين من عرض محتويات مكتبات الصور؛ أولهما : عرض كل مجموعات الصور التي توجد داخل قاعدة البيانات، مع تقسيم هذه المجموعات في أقسام عامة تحوي مجموعات أكثر تخصصًا، وتكون هنا الصور في شكل لقطات صغيرة، على أن يصاحب هذه اللقطات بعض الحقول النصية التي توضح ماهية هذه الصورة وموضوعها ليتمكن الباحث من الحكم على صلة هذه الصورة بموضوع بحثه. ويتناسب هذا الأسلوب في التصفح فقط مع قواعد البيانات ذات الحجم الصغير من الصور؛ لأن قواعد بيانات الصور كبيرة العدد لن يجد المستفيد الوقت الكافي للتعرف إلى كل ما فيها وتحديد ما يناسب احتياجاته الموضوعية. أما الشكل الثاني فهو تكشيف كل مجموعات الصور داخل قاعدة البيانات ثم وضع المصطلحات في الترتيب المنطقي لها، على أن تكون في ترتيب من الأعم إلى الخاص إلى الأكثر تخصصيًا، وربط كل مجموعة من الصور بالمصطلح الذي يعبر عن موضوع هذه الصور باستخدام تقنية الربط الفائق. وهذا من الممكن أن يكون ملائما لطبيعة قواعد بيانات الصور على الويب ذات الأعداد المليونية.

وقد فطنت كل من C. Olivia Frost و Anna Noakes إلى أهمية أسلوب التصفح في استرجاع الصور. وتقدم دراستهما نظامًا يقوم على استرجاع الصور باستخدام التصفح من خلال الاعتماد على نظام التكشيف Art and Architecture Thesaurus يعرف اختصارًا (AAT) ويتم تقسيم المصطلحات إلى قوائم موضوعية Categories تعبر عن مضمون الصور.[١] وهذه الدراسة هي عرض لتجربة قامت بها مدرسة المعلومات بجامعة University of Michigan لإنشاء نظام لاسترجاع الصور المبني على النص وأطلق على هذا النظام اسم SI Art Image Browser. واستكملت هذه الدراسة على

---

(١) Frost, C. Olivia and Noakes, Anna. Browsing Images Using Broad Classification Categories, the University of Michigan, [2000?], visited 1/6/2002, ited at: http://www.dlib.org/dlib/november84/wang/11wang.html

أسئلة بحثية أهمها هل يمكن أن يوفر أسلوب التصفح المبني على نظام تصنيف المصطلحات الموضوعية وسيلة فعالة لاسترجاع الصور؟ وكيف يمكن للمستفيدين أن يحددوا موضع صورة واحدة أو مجموعة من الصور داخل هذا النظام ؟ وما الفروق في الأعداد بين أسلوبي البحث الحر والتصفح؟ واعتمد هذا النظام على واجهة تحوي خياري الاسترجاع إما بالبحث الحر أو بالتصفح، وتم عمل مداخل أساسية للتصفح من خلالها وفق ما يختار الباحث منها وهي مدخل الفنان artist =author، ونوع الوسيط، والموضوع، والتاريخ، والوسط المرئي والعنوان. وعن طريق كل مدخل من هذه المداخل يمكن للباحث أن يتصفح كل مجموعات الصور في قاعدة بيانات النظام.

وذكرت هذه التجربة أن أسلوب الاسترجاع بالتصفح يجعل احتمالات الخطأ في تحديد الصور المطلوبة والمسترجعة في أدنى حد لها، هذا لأن وصف الصورة في الكلمات والمصطلحات المنفردة أو العبارات التي يتم بحثها بالكلمات المفتاحية تتسم بالضعف في التعبير عن موضوع الصورة، فضلا عن أن تلك النظم ترتكز على أساس الاستدعاء العالي مقابل التحقيق المنخفض. كما أن المستخدم الذي يبحث عن صورة بعينها يمكنه اختيار وتحديد مجموعة من الصور التي تلائم موضوعه خلال عملية تصفح كل مجموعة الصور.

وقد ذكرت الدراسة النتائج الأخيرة التي أعطت مؤشرًا على أهمية أسلوب التصفح في تحقيق درجة أعلى من التحقيق لاسترجاع الصور. وأهم النتائج أنه بصفة عامة قد أنهى كل المستخدمين المتمرسين عمليات البحث في وقت قصير مع الوصول إلى أعلى درجة تحقيق.. بينما زاد وقت البحث قليلا لمن هم أقل مهارة في عمليات البحث. ومن بين ٣١ من مستخدمي النظام أكد ٢٢ أن النظام قد حقق ١٠٠٪ من عدد مرات البحث وإخراج النتائج.. بينما أكد اثنان فقط من مستخدمي النظام أن الإصابة في البحث كانت بمقدار ٢٥٪ من مرات البحث، ولم يذكر أي من مستخدمي النظام فشل النظام في الاستدعاء المطلوب. وأظهرت النتائج أهمية وجود أكثر من أسلوب لاسترجاع الصور سواء التصفح أو البحث الحر، حيث يكمل كل من الأسلوبين الآخر. كما ذكر مستخدمو النظام أيضا الأهمية الكبرى للوصف المصاحب للصور، بالإضافة إلى أهمية الترتيب المنطقي للقوائم الموضوعية التي يسرت من عمليات البحث داخل قاعدة البيانات. واختلف مستخدمو

النظام من المتمرسين وغير المتمرسين في البحث ؛ في إجابتهم على سؤال أي المعلومات يريد المستفيد عرضها مع الصورة، حيث أكد ذوو الخبرة أنه يمكن عرض معلومات عن سيرة الفنان صاحب اللوحة أو الصورة، ويمكن أيضا عرض معلومات نقدية خاصة بالصورة ككيان فني، خاصة إذا ما تم استخدام هذه الصور في النواحي التعليمية. أما النوع الآخر من المستخدمين فقد أكدوا الحاجة إلى معلومات عن الفنان وروابط تصل بالمجموعات الأخرى من أعماله وأيضا روابط إلى صور أخرى تشبه خصائص الصورة المنتقاة.

وخرجت الدراسة بنتيجة مؤداها أن نظام الاسترجاع يتأثر بشكل كبير بطبيعة المستخدمين، حيث لم يكن البحث في قوائم نوع الوسيط type طبيعة الوسيط medium ذا قابلية من جانب المستخدمين الأقل مهارة في البحث .

ولقد ذكرت جمعية الخدمة الفنية الاستشارية للصور أن حقول البيانات المرتبطة الواصفة لملفات الصور إنما تتكون أساسا من شقين ؛ الأول : نوع بيانات الوصف أو بنيتها الأساسية وهي كحقول بيانات مارك أو معيار دبلن كور. الثاني : الكلمات التي ستوضع داخل الحقول الواصفة لملف الصورة مثل ماهية كلمات وصف المحتوى وطبيعة قائمة رؤوس الموضوعات أو المكنز المستخدم. وعلى ذلك تبرز الأهمية الكبيرة في اختيار أداة ضبط الكلمات الواصفة لمحتوى أوعية الصور الرقمية إذا ما تم الاتفاق على التعامل مع محتوى الصورة باللغة المقيدة أو المضبوطة. وقد قدمت أيضا مجموعة من الاعتبارات الواجب مراعاتها عند اختيار قائمة الضبط في نظام استرجاع الصور الرقمية. وهي كالتالي:-[١]

المستفيدون : تحديد مدى ملاءمة قائمة الموضوعات المختارة للمستفيدين في إعطاء المعنى الكامل للتعبير عن المحتوى.

طبيعة المجموعات : حيث أن المجموعات قليلة العدد لا تحتاج إلى أداة تحتوي على تفاصيل كثيرة، والعكس صحيح من ذلك.

---

(١) Controlling your language - links to metadata vocabularies. TASI Technical Advisory Service for Images, 2003, visited at 12/12/2003, visited at http://www.tasi.ac.uk/resources/vocabs.html

مهارة اختصاصيي المعلومات : وهي مدى ملاءمة طبيعة أداة ضبط المصطلحات لمهارة العاملين داخل نظام الاسترجاع، فلا يجب الاعتماد على الأدوات المعقدة التي لا تناسب مهارة العاملين في النظام.

مجتمع الخدمة : حيث إنه من المناسب الاعتماد على أداة ضبط لمصطلحات تلائم المجموعات الأكثر استخدامًا من جانب العاملين.

إضافة إلى ذلك فهناك مجموعة من العوامل المساعدة التي يجب أن تؤخذ في الاعتبار عند الأخذ بأسلوب التصفح مثل سرعة تحميل الصور على شاشة عرض النظام ؛ حيث يتوقف عل ذلك الوقت الذي يمكن فيه للمستفيد تصفح كامل قاعدة البيانات، سواء طال هذا الوقت أم قصر، نتيجة لأداء نظام الاسترجاع. وكذلك حجم شاشة العرض وعدد الوحدات المسترجعة في كل شاشة عرض، بجانب ذلك أيضا حجم كل وحدة من الصور بحيث تكون الوحدة على أصغر حد لها بما لا يتعارض مع إمكانية إظهار تفاصيلها أمام الباحث. ويتعلق بهذا الأمر الجودة التي تم بها إدخال الصورة إلى النظام. ومن ذلك أيضا تصميم واجهة الاستخدام للنظام ومداخل التصفح أمام الباحث، بما يوفر للباحث مداخل غير معقدة للبحث ونظام تصنيف المداخل بما يسمح للباحث أن يتحرك من الأعم إلى الأخص بمرونة.

٣/١/٢/١/١ علاقة التصفح باللغة المضبوطة

يعتمد التكشيف على اختيار الكلمات المفتاحية التي تصف محتوى وسيط المعلومات، حيث يعد إحدى عمليات التحليل الموضوعي التي تقف عند مستوى المحتوى الموضوعي والتعبير عنه بالكلمات أو المصطلحات المفتاحية، وهو خلاف ما يحدث في عمليات التحليل الأخرى مثل التصنيف في التعبير عن الموضوع بواسطة رموز من حروف أو أرقام، وعملية الاستخلاص التي تتعدى مستوى الكلمات إلى مستوى جمل واصفة مكونة شكل المستخلص النهائي. وبرز مصطلح لغة التكشيف من خلال طبيعة المصطلحات المستخدمة للتعبير عن المحتوى الموضوعي ؛ فإذا ما تم الاعتماد على قائمة من المصطلحات المعدة والمختارة مسبقا، أصبحت اللغة مقيدة أو محكومة أو مضبوطة. أما إذا تم الاعتماد على طبيعة الكلمات التي وردت في النص أو العنوان وفقًا لما اختاره المؤلف، أصبحت اللغة طبيعية. ويتم الاعتماد فقط في هذه الحالة على طبيعة آليات البحث داخل

قاعدة البيانات وليس على المطابقة بين قائمة المصطلحات والمداخل التي وضعت أسفلها مصادر المعلومات داخل قاعدة البيانات.

ويمثل التصفح برؤوس الموضوعات أو المصطلحات داخل الويب نموذجًا من نماذج الاعتماد على اللغة المقيدة في التكشيف وتحليل محتوى مصادر الويب، ذلك لما يرتبط به من تقيد مستخدم نظام الاسترجاع بقائمة مصطلحات هرمية التنظيم يتم ربطها بمجموعة مصادر المعلومات التي تتوافق مع دلالة كل من المصطلحات. وهناك من أدوات بحث الويب مثل Yahoo ما يعتمد على قائمة من المصطلحات الخاصة بالموقع تندرج أسفل منها صفحات الويب التي تلبي الحاجة الموضوعية الخاصة بكل من المصطلحات. ويمكن في هذه الحالة الاعتماد على أي من المكانز الخاصة بالتكشيف في مجال محدد أو مجالات متعددة بصياغة شكل من أشكال التصفح الداعم للتحكم في مجموعة ثابتة من المصطلحات تتكون على أثرها قوائم رئيسة ثم فرعية تدرج أسفل منها محتويات قاعدة البيانات ككل. وقد اعتمدت بعض أدوات البحث على نظم تصنيف معدة مسبقًا مثل تصنيف ديوي والتصنيف العشري العالمي، أو أن تعتمد بعض المواقع الأخرى على صياغة أشكال من نظم التصنيف تلائم المحتويات الخاصة بها من مصادر الويب. وقد ذكر هانئ عطية بعضًا من جهود تصنيف المعلومات على الويب في سرد مجموعة من نظم التصنيف المعدة مسبقًا أو ذاتية البناء، بالإضافة التي تقسيم تلك النظم إلى أنماط عامة متعددة مثل نظم التصنيف المكتبية كنظام مكتبة الكونجرس، ونظم التصنيف القومية كالتصنيف الأساس في هولندا والتصنيف القومي السويدي للمكتبات العامة، ونظم التصنيف الموضوعية كالتصنيف الخاص بمجال الحاسب الآلي والتصنيف الموضوعي للرياضيات، ونظم التصنيف المولدة في البيئة الرقمية كتصنيف بوابة Edinburgh Engineering Virtual Library وتصنيف بوابة Social Science Information Gateway. وقد أكد الكاتب وجود ١٦ تصنيفًا على الأقل تخدم تنظيم المعلومات على الويب ثلاثة منها عالمية، واثنان قوميان، وأربعة موضوعية، وسبعة خاصة، مما يعني البحث الدؤوب حول نظام تصنيف يوفر الدرجة الأفضل من التحقيق داخل النتائج المسترجعة. ومن بين ما أكدت عليه هذه الدراسة أيضًا أن أغلب المواقع قد لجأت إلى استخدام نظم التصنيف العالمية خلاف النظم القومية، وذلك لما لها من ثبات وتجربة سابقة في الاستخدام لاسترجاع المعلومات. كما جاء

أيضًا من بين النتائج أن أغلب المواقع قد لجأ إلى استخدام المكانز بدلا من قوائم رؤوس الموضوعات لما لها من سمات مهمة مثل الاعتماد على المصطلحات المفردة إضافة إلى التخصص الموضوعي في مجالات محددة يغطيها مجال موقع الويب فقط.(١)

ويمثل العرض السابق المحاولات التي تسعى إليها أدوات بحث الويب لتحسين الاسترجاع المعتمد على استخدام اللغات المقيدة من نظم التصنيف والتكشيف العالمية أو ذاتية البناء. وتستخدم فقط أدلة بحث الويب قوائم المصطلحات هرمية البناء لاسترجاع المعلومات بأسلوب التصفح الذي تفتقد إليه محركات البحث المعتمدة على أسلوب البحث بالكلمات الحرة. وعلى ذلك فإن كلًا من قوائم مصطلحات التصفح وأدلة البحث وجهان لعملة واحدة هي لغات التكشيف المقيدة أو المضبوطة؛ إلا أن محركات البحث الآن وعلى رأسها Google قد عملت على تدعيم عمليات البحث الحر لديها بمجوعة من قوائم المصطلحات هرمية البناء ذاتية التكوين للوصول إلى مستوى أعلى من التحقيق في نتائج بحث قاعدة البيانات لديها.

٢/١/٢/١/٣ مشكلات التصفح في استرجاع الصور الرقمية

تقف مشكلات التكشيف بقوائم التصفح عند مستوى اختيار اللفظ أو المصطلح الملائم للوضع في قائمة الموقع والذي سوف تندرج أسفل منه مجموعة مصادر المعلومات الرقمية، إن السمة الأساسية في بناء قوائم رؤوس الموضوعات هي اعتماد مجموعة من المصطلحات دون الأخرى التي تقع معها في نطاق الترادف للتعبير عن الموضوعات ومن ثم تنظيم مصادر المعلومات أسفل هذه المصطلحات للبحث بها من جانب مستخدم الويب، ويعتمد اختيار رأس الموضوع و الواصفة في المكانز على أسس محددة مثل مدى شيوع هذا المصطلح بين لمستخدمين، واختيار رأس الموضوع المخصص، والصيغة العامة للموضوع.

وعلى ذلك فإنه يمكن القول إن تخصيص قائمة رؤوس موضوعات بعينها داخل نظام استرجاع الصور لبناء أسلوب التصفح يعتمد في البدلة على دراسة

(١) هانئ محيى الدين عطية. جهود تصنيف المعرفة في البيئة الإلكترونية: دراسة استكشافية -.**الأكاديمية للمكتبات والوثائق والنظم المعلوماتية**، ع١ (٢٠٠٤)-. ص٥-٧٧.

مجموعات الصور داخل قاعدة البيانات، كما يعتمد على دراسة محتوى الصور الرقمية وتعدد الموضوعات التي تحتويها، فضلا عن تعدد المصطلحات التي تندرج أسفل منها الصورة الرقمية لما تتميز به من وصف موضوعي ذات أبعاد متباينة. إن اختيار المصطلح الكشفي من جانب المكشف يعتمد على المعنى الملائم لموضوع الصورة ومن ثم يمكن هنا تلافي مشكلات معنى المصطلحات الكشفية للموضوع؛ إلا أن اختيار اللفظ الملائم هو ما يؤثر فعليًا على مستويات التحقيق والاستدعاء في استرجاع مصادر المعلومات الرقمية بصفة عامة والصور الرقمية بصفة خاصة. ويكمن هذا خاصة عند العمل مع اللغة العربية التي تبرز مشكلاتها في المعنى واللفظ عند العمل باللغة الطبيعية، وهو ما يسمى البحث الحر بالكلمات المفتاحية المعتمد على التكشيف الذاتي لمحركات بحث الويب.

٣/١/٢/٢ البحث الحر بالكلمات المفتاحية

ينقسم البحث الحر بالكلمات المفتاحية إما بالواصفات المنفردة أو بالواصفات المتعددة ولكل منها خصائصها. وفيما يلي وصف كل منها بالتفصيل:-

٣/١/٢/٢/١ الوصف الموضوعي بالواصفات المنفردة

يعتمد وصف محتوى الصور في هذه الحالة على مجموعة من المكانز ورؤوس الموضوعات. وتعتمد مكتبات الصور في وصفها الموضوعي لمجموعاتها على مكانز متخصصة للمواد المصورة أو قوائم رؤوس الموضوعات العامة. والأمثلة هنا كثيرة لهذه الأدوات مثل مكنز الفنون والعمارة Art and Architecture Thesaurus AAT ويحتوي هذا المكنز على ١٢٠ ألف مصطلح تصف مختلف وسائط المعلومات الثابتة، ومكنز موضوعات مكتبة الكونجرس للمواد المرسومة Library of Congress Thesaurus for Graphic Materials ويعرف اختصارًا (LCTGM)، وأيضا مكنز الأسماء الجغرافية Thesaurus of Geographic Names ويعرف اختصارًا (TGN). وبالنظر إلى المسميات السابقة لأدوات تكشيف الصور والمواد المرئية، فإنها جميعًا تعطي مفهومًا مختلفًا عن المفهوم التقليدي للمكنز، حيث تعكس هذه الأدوات جانب التخصص في المكنز بشكل آخر، فالتخصص لهذه المكانز ليس من حيث الموضوع وإنما من حيث شكل المادة الموصوفة.

ولقد تناول تقرير المعهد العلمي للمعلومات المصورة Institute for Image Data research

تكشيف الصور بالكلمات المفتاحية، وأظهر المعهد في هذا التقرير أن التكشيف بالكلمات المفتاحية لا يناسب طبيعة المستفيدين في التعرف إلى احتياجاتهم من الصور، كما أنه لا يناسب تحليل محتوى هذا النوع من الأوعية الذي يتضمن بداخله الكثير من الموضوعات التي يمكن أن تفيد أكثر من صنف من الباحثين في مجالات مختلفة. وقد ذكر التقرير أن هذا الأمر يرجع إلى عاملين أولهما: أن الوقت المستغرق في عملية تكشيف وتحليل وحدة من الصور تستغرق ما بين ٧ دقائق للصور الفوتوغرافية إلى ٤٠ دقيقة لإحدى الشفافيات وذلك اعتمادًا على أداة AAT. ثانيهما: أنه بمراجعة الكلمات المفتاحية التي وضعها المكشفون للتعبير عن المحتوى كانت تختلف في مفردات كثيرة عن تلك التي وضعها المستفيدون لوصف نفس المجموعة من الصور .[١]

والجدير بالذكر أن ذكر جوانب السلب السابقة لا تنقص من أهمية التكشيف للصور، وإنما العمل هنا لم يكن متكامل الجوانب ؛ فنظام الاسترجاع له جوانب عدة ليس فقط الكلمات المفتاحية، فقد يساعد على تحقيق درجة أكبر من الدقة أن توضع قوائم استناد للمصطلحات تساعد الباحثين في التعبير الصحيح عن الموضوع، كما يمكن ربط الصور المتقاربة في المحتوى معًا من خلال الربط الفائق تساعد هذه الصور المرتبطة معًا في تحديد المجموعة التي يسعى إليها الباحث. ومن جانب آخر، فان النتائج التي توصل إليها هذا التقرير تؤكد على أن الصعوبة الأولى في عملية تحليل محتوى المواد المصورة تأتي في التعرف إلى كل الجوانب الموضوعية التي تغطيها الصور. كما أن النتائج الخاصة بتكشيف الكلمات المفتاحية تؤكد أيضًا مدى أهمية وصف المحتوى بالشكل السردي في شكل ملخص يحكي محتوى الصور، ذلك لأن هذا الشكل يمكن أن يكون أقرب تمثيلا إلى ذهن مستخدمي نظام استرجاع الصور.

ويمكن القول إن هذا الأسلوب من أساليب الاسترجاع للصور وهو استخدام النص ليس على الدرجة الكاملة التي تحقق الدقة المطلوبة في استرجاع الصور. وهذا يرجع لأن كل موضوعات الصورة التي توجد بداخلها لا يتم تمثيلها بالشكل الذي يعكس كل محتويات الصورة أو يكون بالشكل الخاطئ في تعريف الموضوعات.

---

(١) Eakins, John P and Graham, Margaret E, Optc. P25.

وتقوم الكلمات المفتاحية بالدور الرئيس في عرض كل الموضوعات. وتفسير هذه المشكلة يكمن في تقسيم التحليل الموضوعي للصورة إلى مستويات ثلاثة، ويكون وصف موضوعات الصورة على النحو التالي:-

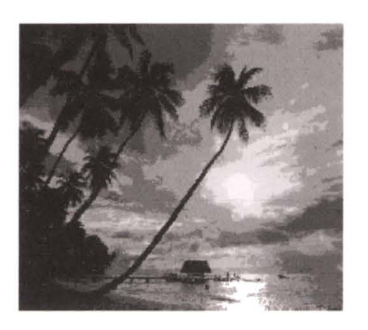
المستوى الأول : يختص هذا المستوى بالعناصر التي تقع عليها العين أول ما ترى الصورة ويستخدم المكشف المكنز أو أداة التكشيف في التعبير عن كل عنصر بما يوافقه من واصفات في المكنز أو قائمة رؤوس الموضوعات. ويجب أن يكون مبدأ الإحاطة هو أساس العمل في تكشيف الصورة، وذلك لعاملين: الأول أنه في الوقت الذي على المكشف أن يصف كل موضوعات الصورة الرئيسة، فإن الكلمات المنفردة لا تعطي المكشف الإيجاز في التعبير. وهي ما يمكن أن نطلق عليها العناصر الحاكمة للمكشف.

المستوى الثاني: يكون لإبراز العناصر الثانوية التي تعتمد عليها الصورة في استكمال كل أجزاء الصورة، ولا تقل هذه العناصر أهمية عن العناصر الأولى ؛ ذلك لأن الاسترجاع بالكلمات المفتاحية دائما ما يعتمد على مصطلحين أو أكثر لتحديد موضوع البحث. وهنا يتم الحديث عن مختلف الأشياء الأخرى مهما بلغ صغر هذه العناصر. وهي ما يمكن أن نطلق عليها العناصر الذهنية.

المستوى الثالث: يعبر هذا المستوى عن العناصر المعنوية التي تريد الصورة التعبير عنها وإيصالها إلى من يراها، بجانب العناصر الأخرى. ويعد هذا المستوى أصعب المستويات الثلاثة أداءً، ذلك لأن الاختلاف بين المستفيدين والمكشفين في الصور يقع في هذا المستوى؛ لأن المعنى الموجود في الصورة سوف يختلف من شخص إلى آخر، من هنا فإنه من الأهمية بمكان أن يتم تكشيف الصور مع تمام الوعي بنوع المستفيدين من نظام استرجاع الصور. وهي ما يمكن أن نطلق عليها العناصر الضمنية أو المعنوية.

وتوضح الصورتان رقم (١) ورقم (٢) مثالين لبيان مستويات تكشيف الصور، واختار الباحث نوعين من أنواع الصور ؛ فتعبر الصورة رقم (١) عن الصور الفنية أو الطبيعية، بينما تعبر الصورة رقم (٢) عن الصور الإعلامية. ويتم تكشيف الصور على المستويات الثلاثة :-

مثال لتحليل الصورة بالكلمات المفتاحية



صورة مكشفة رقم ١





صورة مكشفة رقم ٢

الشكل رقم (٣-٢) يوضح نموذجي تكشيف الصور الرقمية اطبيعية

ويمكن التعبير عن المستويات الثلاثة للتكشيف في هاتين الصورتن كما يلي:-

	المستوى الأول	المستوى الثاني	المستوى الثالث
الصورة الأولى	النخيل	الرمال	الطبيعة
	المياه	السحاب	الغروب
	الشاطئ	الشمس	الهدوء
	الشجر		الجمال
	البحر		الحزن
الصورة الثانية	يانج زيمين	الزهور	التفاهم
	فلاديمير بوتين	ستارة	التوافق
	الصين	علم الصين	الصداقة
	روسيا	منضدة	التعاون
			السياسة

إن تكامل هذه الواصفات معًا على اختلاف مستويات التكيف يؤدي إلى درجة عالية من الدقة في حال التعامل الجيد مع عناصر الصورة ككل. ومن جهة أخرى يمكن التحكم في كل من الإحاطة والتخصيص كمبادئ أسسية للتكشيف، بتكوين الواصفات التي تعبر عن ذلك.

٢/٢/٢/١/٣ الوصف الموضوعي بالعبارات

يختلف الوصف الموضوعي بالعبارات عن الوصف بالكلمات المفتاحية، فالأخير يلزم المكشف بالمرور على المستويات الثلاثة في التكشيف. أما الوصف الموضوعي بالعبارات فإنه يمكن دمج المستويات المختلفة للتكشيف في العبارات، لتصبح مستوى واحدًا. ويتحدد مستوى التكشيف هنا بعدد العبارات الشارحة لمحتوى الصورة، سواء كان مفصلًا فيزداد فيه عدد العبارات أو مختصرًا فيقل على أثره عدد العبارات الواصفة. وتبعًا للمثال السابق يكون الوصف بالعبارات المنفصلة كما يلي:-

	الوصف بالعبارات
الصورة الأولى	مجموعة أشجار نخيل شمس وراء السحاب الناس والشاطئ مياه البحر بجانب الرمال جمال رؤية البحر هدوء وقت الغروب حزن اختفاء الشمس
الصورة الثانية	اجتماع الرئيس بوتين مع الرئيس زيمين علم الصين الشعبية العلاقة بين روسيا والصين التعاون بين الدول الكبرى نتاج السلام الدولي العلاقات الدولية

لقد اعتمد الباحث في صياغة العبارات السابقة على إبراز جوانب المحتوى المادي أو الظاهرة أمام العين ثم الجوانب غير المرئية. ويختلف مستوى التكشيف من نوع إلى آخر في الصور الرقمية، إلى جانب اختلاف عدد الكلمات المفتاحية لوصف محتوى الصورة من نوع إلى آخر في كلام من الصور العلمية والإعلامية والطبيعية.

على الجانب الآخر، فإن كلًّا من نسبة الاستدعاء والتحقيق تتأثران بعدد العبارات المستخدمة في وصف محتوى الصورة، ذلك لأنه كلما زادت عبارات

الوصف المستخدمة زادت نسبة الاستدعاء في هذا النظام، وعكس ذلك هو ما يحدث في نقص عبارات الوصف. ذلك لأن الزيادة في عبارات الوصف دائما ما تحمل معها الكثير من الكلمات المساعدة في الوصف التي قد لا تقترب لموضوع الصورة الرئيس؛ إلا أنها قد تدخل في إستراتيجيات بحث المستفيد داخل محرك البحث مما ينتج عنه استدعاء لمجموعات من الصور مختلفة في الموضوع الرئيس متفقة في الكلمات المساعدة لوصف الصورة.

٣/٢/٢/١/٣ الوصف الموضوعي بالجمل المترابطة

يعتمد هذا النوع من وصف محتوى وموضوعات الصور على النص الشارح في شكل ملخص summary. على أن يكون هذا الملخص بنيانًا كاملًا لكل ما تعنيه الصورة. ويمكن أن يكون هذا الأسلوب هو الأكثر بلاغة في التعبير عن محتوى الصورة، كما أنه يمكن أن يكون الأجدى في استرجاع الصور اعتمادًا على النص، هذا إنما يرجع إلى عوامل كثيرة مثل طبيعة المستفيد التي قد تميل إلى شرح الموضوع أثناء عملية البحث، حتى أن الكثير من المستخدمين لمحركات البحث لا يستخدمون الكلمات المفتاحية، بل يستخدمون الجمل والعبارات. كذلك عناصر الصورة وهي أركان التكشيف التي تحوي فيما بينها الكثير من المعاني واجبة الذكر والتي يمكن أن يكون لها أكبر الأثر في استرجاع صور بعينها. وتمثيلا على الصور السابقة، فإن وصف المحتوى باستخدام الجمل يمكن أن يكون على الشكل التالي:-

	الوصف بالجمل المترابطة
الصورة الأولى	مجموعة من أشجار النخيل تمتد إلى قلب السحاب وتلتقي أمواج مياه البحر مع الرمال مكونة الشاطئ، وعلى الجانب الآخر تختفي الشمس الصفراء وراء السحاب معلنة وقت الغروب بما له من هدوء وحزن وجمال
الصورة الثانية	الرئيس الروسي فلاديمير بوتين والرئيس الصيني يانج زيمين في اجتماع يجمعهما، تعكس الصورة العلاقة بين دولة روسيا الاتحادية ودولة الصين، وتعبر جوانب الزهور والعلم الصيني عن الصداقة والتعاون وجوانب المشاورة بين البلدين، والصورة من نتاج السلام والتفاهم الدولي

إن ما سبق يمثل المستوى الثالث من مستويات التحليل الموضوعي للصور، يذكر البعد المعنوي للصورة، حيث اعتمد الباحث على استخدام الكلمات المفتاحية في المستوى الأول، والعبارات المفتاحية في المستوى الثاني.

لقد أصبحت الفهرسة الوصفية والتحليل الموضوعي للصور هما السبيل الوحيد أمام قواعد بيانات الصور ومحركات البحث لتفادي مشكلات جمة في استرجاع الصور. فيذكر Greg Notess[1] في حديثه عن بحث ملفات الصور على الويب، إن بحث واسترجاع ملفات الصور، خاصة عند البحث عنه من جانب محركات البحث، دائما ما كان يعتمد على النص المصاحب للصورة. وهذا النص يتمثل في الكلمات التي تكون اسم ملف الصورة وامتداده أو أن تكون الكلمات المعدودة المصاحبة للصورة أو تكون تلك التي تكون الرموز المعيارية Meta Tags. ولم تكن هذه الكلمات المحدودة التي لا تعبر عن كل موضوعات الصورة مناسبة لاسترجاع أمثل لمادة الصور. ومن ثم كان الحل في الرجوع إلى قواعد الفهرسة الوصفية، واستخدام حقول الوصف الملائمة للصورة مثل منتج الصورة والتاريخ واللون واسم الملف... الخ، مع إدخال الوصف الموضوعي المساعد في الاسترجاع. ولقد كانت حقول الوصوف الببليوجرافي للصور والتحليل الموضوعي لها، هو ما أوصى به مصممو قواعد البيانات وعلوم الحاسب حيث أشار Joe Farace أن حقول الوصف للصور والكلمات الكشفية المصاحبة لها هما أهم جوانب تصميم قواعد برامج قواعد بيانات الصورة الجيدة، التي يمكن أن تخدم الباحثين بما ينتظرونه من نظم استرجاع الصور[2]. والجوانب الواجب مراعاتها في تصميم قواعد بيانات الصور كما يلي :-

- ✓ توفير برامج تستطيع قراءة كل أنواع الملفات التي يمكن أن توجد عليها الصور، والتي يبلغ عددها أكثر من ١٠٠ نوع، القليل منها يستخدم للصور.
- ✓ توفير برنامج قاعدة البيانات بعرض ملفات الصور في لقطات أثناء عرض النتائج، وإذا لم توجد هذه الخاصية، فان برامج مثل ThumbsPlus يستطيع إنتاج هذه اللقطات من ملفات الصور.
- ✓ توفير قاعدة بيانات جيدة تتيح للباحثين البحث بالكلمات المفتاحية المرافقة لملف الصور والتي تكون في أقل حالاتها عند استخدام كلمة أو كلمتين كواصفتين.

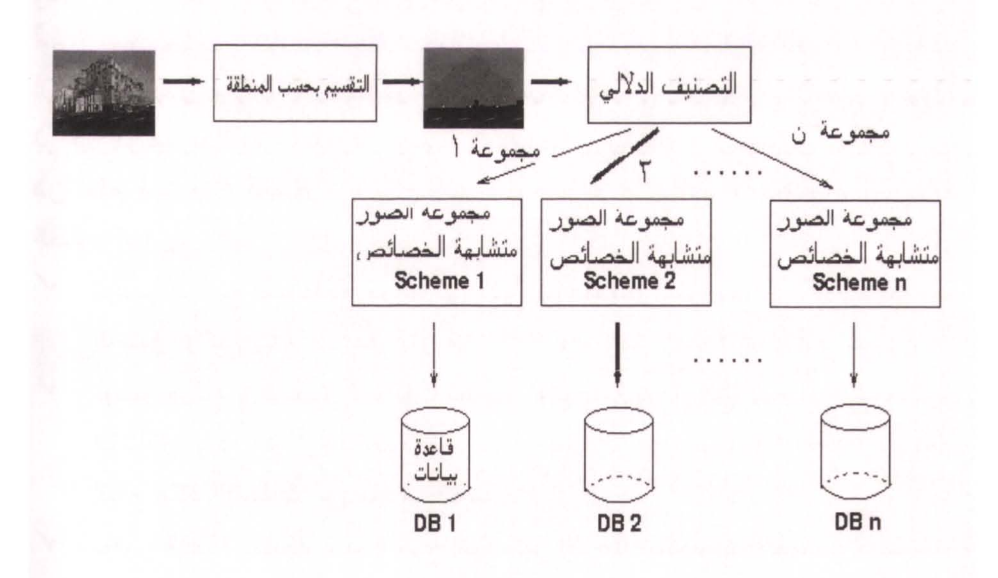
---

(١) Notess, Greg. Searching beyond text: issues with multimedia searching, information today, 2000, visited 27/9/2002,: http://www.onlinemag.net/OL2000/net9.html

(٢) Farace, Joe,. Take Control of Your Image Assets, computer user, visited at 21/9/2002, 2000, http://www.computeruser.com/articles/1907,2,3,1,0701,00.html

✓ توفير برنامج استرجاع صور الكفء يضم حقولًا واصفة تمثل تفاصيل الصور وسبيل استرجاعها.

وقد عملت بعض الدراسات مثل دراسة James Ze Wang[(١)] على استخدام تكشيف الصور الرقمية في تصنيف مجموعات الصور حسب الصفة الدلالية لكل منها؛ حيث يمكن استخدام مجموعة من رؤوس الموضوعات لإقامة مجموعات رئيسة وفرعية في التعامل مع موضوعات الصور الرقمية، ومن ثم تشكيل مجموعات مختلفة من الصور تتفق كل منها داخليًا في الخصائص الدلالية مع اختلافها مع المجموعات الأخرى بما يمكن معه تيسير البحث والاسترجاع داخل قاعدة بيانات الصور. وقد اختصت هذه الدراسة بمجموعات الصور المكشفة والمصنفة تبعًا للمنطقة أو الإقليم التابع له محتوى الصورة. ويوضح الشكل رقم (٣-٣) المنهج المعتمد في تنظيم واسترجاع الصور الرقمية المكشفة :-



الشكل رقم (٣-٣)

يوضح تصنيف الصور المبني على التكشيف الدلالي للمجموعات

---

(١) Wang, James Ze ...[etc]. Semantics-sensitive Retrieval for Digital Picture Libraries, D-Lib Magazine, 1999, visited at 12/10/2003, cited http://www.dlib.org/dlib/november99/wang/11wang.html

٢/٣ معالجة اللغة العربية في نظام استرجاع الصور

تتسم اللغة العربية بالكثير من السمات التي تنفرد بها عن غيرها من اللغات الأخرى، وعلى ذلك فإن ثمة تأثيرًا يمكن أن تحدثه اللغة داخل نظام الاسترجاع، ويجلو أثره أكثر على التحقيق والاستدعاء داخل نظام استرجاع الصور باللغة العربية. ويختص هذا الجزء من الدراسة باستيضاح خصائص اللغة العربية وكيفية تطويعها بشكل يحقق الكفاية لنظام استرجاع صور يعتمد في بنيته على اللغة العربية. والكيفية التي تكون عليها مصطلحات اللغة العربية داعمة مقياسي الاستدعاء والتحقيق، بما يرجوه المستفيد من نظام الاسترجاع.

١/٢/٣ مشكلات البحث الحر باللغة العربية في البيئة الرقمية

ينطوي مصطلح البحث الحر على استخدام البحث باللغة الطبيعية داخل النصوص الكاملة لمصادر المعلومات. وتلتقي دراسة اللغة الطبيعية باسترجاع الصور عند دراسة الشكل الأخير من الوصف الموضوعي للصورة، وهو الوصف في شكل جمل سردية أو نص شارح، فهو أقرب أشكال التحليل الموضوعي إلى مستخدمي الصور. وينتج عن العمل باللغة الطبيعية في بحث المصادر العربية الكثير من المشكلات التي تعكس خصائص اللغة العربية و يمكن حصرها في الفئات التالية:-

١/١/٢/٣ الاسترجاع اللفظي

وهي مشكلات قد قام بحلها وجود برنامج التحليل الصرفي المضاف إلى معظم محركات بحث، ولكنها ستظل قائمة لو لم يزود محرك البحث ببرنامج التحليل الصرفي. ومن ثم ينوه عنها الباحث هنا للتعريف بأنواع هذه المشكلات المرتبطة بخصائص اللغة العربية في حالة غياب المحلل الصرفي، وهي:-

١/١/١/٢/٣ صيغة الفعل

وهي الصيغ المختلفة لمشتقات الفعل، فالبحث عن كلمة (الحج) مع تحديد خيار البحث بالمشتقات، يستدعي وحدات تتضمن (الحج، حجة، حاج، يحج... إلخ).

٢/١/١/٢/٣ الإملاء ورسم الكلمات

وهي المشكلات الناجمة عن الأشكال المختلفة لكتابة الكلمات خاصة في

حروف الهاء الختامية للكلمة مع التاء المربوطة، والياء اللازمة والألف المقصورة المرسومة بالياء، وهمزة القطع والوصل. وعلى ذلك فان بحثا لكلمة (احمد) سوف يستدعي أيضًا الشكل الآخر للكلمة (أحمد).

٣/١/١/٢/٣ الحروف والأدوات "قائمة التوقف"

وهي التمييز بين الحروف والأدوات التي قد تعيق البحث مثل التمييز بين أل التعريف باعتبارها لواصق في الكلمة أو كأصل فكلمة مثل "الله" تعد فيها (ألـ) أصلية بينما كلمة مثل (الكتاب) لا تعد كذلك. وتعمل نظم الاسترجاع على حذف (ألـ) التعريف عادة في مقدمة الكلمات ولكن ذلك يتطلب وجود قائمة كلمات التوقف التي تستبعد بعض الكلمات ومن ثم يكون من المناسب هنا وضع الكلمات التي تأتي فيها ألـ أصلية مثل (الذي" و"التي) وغيرها.

٢/١/٢/٣ الاسترجاع بالمعنى

وهي المشكلات التي تتعلق بالكلمات المفتاحية المستخدمة في وصف الصور الرقمية، ومن ثم فهي تتعدى الحدود اللفظية للمعنى الضمني. ومثل هذه المشكلات تعمل بعض محركات البحث على إيجاد حلول لها؛ إلا أنها غير قاطعة، ومن أهم هذه المشكلات:-

١/٢/١/٢/٣ المترادفات

يعرف الترادف على أنه "توالي الألفاظ المفردة الدالة على مسمى واحد باعتبار معنى واحد فيخرج عن هذا دلالة اللفظين على مسمى واحد، لا باعتبار واحد، بل باعتبار صفتين كالصارم والمهند أو باعتبار الصفة وصفة الصفة كالفصيح والناطق "[(١)]. وأضاف عبد الرحمن المحسني إلى ذلك أن ثمة قضية لغوية تتعلق بالترادف وهي إنكار الترادف وإثباته، فقد انقسم أهل اللغة إلى فريقين في هذا الشان، ويرى فريق الإنكار أن الشارع حكيم ولا توجد كلمة إلا ولها دلالة خاصة بها، وتبعًا لذلك فلا ترادف في اللغة. أما فريق الإثبات يرى أن الترادف موجود باللغة وأن هناك كلمات عربية عدة مترادفة لم تأت عبثا وإنما هناك

---

(١) عبد الرحمن بن حسن المحسني .**قضية الترادف: النظرية والتطبيق**، تاريخ الاستشهاد ٢٠٠٢/١٠/٢٣، مسار http://www.adabiabha.com/biader/4.htm

مقاصد محددة لذلك، وأكد هؤلاء رأيهم بالأدلة العقلية والآيات القرآنية الكريمة.[١]

أما من منظور نظم استرجاع المعلومات، فإن الترادف يسهم في ضياع جزء من التحقيق للمستفيد في استرجاع المعلومات، على عكس ما يكون الأمر عليه في مشكلة الاشتراك اللفظي حيث يعاني المستفيد من ارتفاع نسبة التشويش (الشوشرة) في النتائج المسترجعة. وعلى الرغم من أن المشكلة لدى الصوينع توقفت عند الكشافات اليدوية وتأثير الترادف على تشتت المداخل في الكشاف، إلا أن البيئة الرقمية لنظم الاسترجاع أعادت صياغة مشكلات اللغة العربية بمستوى بحث آخر خلاف ما يكون في البيئة التقليدية. وقد كانت الإحالات هي الحل الأمثل لمثل هذا النوع من مشكلات الاسترجاع اليدوي لمعلومات اللغة العربية .[٢]

ويمكن تصوير الترادف في مثاله الشهير عن الكلمات الثلاثة المستخدمة لوسيلة الاتصال بالهاتف وهو (الهاتف - المسرة - التليفون) . إن بحثا في محركات البحث التقليدية من شأنه أن يتسبب في ضياع جزء من وحدات المعلومات المنتمية للموضوع.

٢/٢/١/٢/٣ الاشتراك اللفظي

يقصد بالألفاظ المشتركة أو المشترك اللفظي مجموعة من المعاني والمفاهيم المختلفة التي تتوحد في مجموعة حروف واحدة مكونة كلمة ذات شكل واحد. بهذا فإن الباحث عند استخدامه لأحد الكلمات المتسمة بهذه السمة سوف يكون أمام فيض من المعلومات التي لا حاجة له بها، ومن أمثلة ذلك أن يقوم باحث بالبحث عن صور عن (قطر) فيدخل كلمة (قطر) فتأتي "أشكال رياضية تمثل (قطر) الدائرة. ولقد أثمر جهد نظم الاسترجاع بالاعتماد على تقنية التحليل المعجمي آلية البحث بالتشكيل، وعلى الباحث عند إدخال كلمة البحث ثم تحديد خيار البحث بالتشكيل لتعيين المعنى المطلوب اعتمادًا على شكل الكلمة؛ إلا أن قواعد البيانات التي تعنى بالصور لا تقوم بتشكيل الكلمات المفتاحية، ولم يظهر في

---

(١) المصدر السابق نفسه.

(٢) علي السليمان الصوينع .**استرجاع المعلومات في اللغة العربية**.- الرياض: مكتبة الملك فهد الوطنية، ١٩٩٤. ص- ص ٦٨-٧٧

الأفق حتى الآن التقنيات التي تسهم في إصلاح عمليات البحث في مصادر اللغة العربية.

٣/١/٢/٣ اللغة والثقافة

تشكل الثقافة و البيئة التي يتم تكشيف الصور بها عاملا مهما في اختيار الكلمات المفتاحية عن محتوى هذه الصورة. ويمكن لصورة واحدة أن تستخدم لإيصال أكثر من معنى، وقد تكون هذه المعاني ليست فقط في حالة تباين وإنما في حالة تضاد في المعنى ؛ على سبيل المثال فإن صورة واحدة لطفل فلسطيني يحمل حجرًا يرمي به جنديا صهيونيا، يمكن استخدامها لإيصال معنى النضال، كما تستخدم على الجانب الآخر في الإعلام الغربي لإيصال معنى الإرهاب، وهو البعد الثالث في التعامل مع تحليل الصور. وقد ناقش هانئ محيى الدين الأبعاد المختلفة للصورة بمناقشة الطلاب لموضوع الصورة المكشفة في تعلم قراءة الصور وتعليمها.[١] وعلى ذلك فإن معرفة البيئة التي تنتمي لها الصورة له أكبر التأثير في التعامل مع وصف وتحليل الصور الرقمية على الويب التي تمتزج داخلها صور عدة لبيئات متنوعة.

٢/٢/٣ آليات التعامل مع اللغة العربية في البيئة الرقمية

إن حلقة الكلمات المفتاحية والواصفات في نظم استرجاع المعلومات بصفة عامة ونظم استرجاع الصور بصفة خاصة، ترتبط بحلقة أخرى لا تنفك عنها أبدًا هي حلقة اللغة، والمقصود هنا لغة نظام الاسترجاع وملفات الصور، وليس لغة التكشيف.

١/٢/٢/٣ آليات البحث بالكلمة الواحدة

يعد التحليل الصرفي لب عمليات معالجة اللغة الطبيعية، حيث تقوم نظم الاسترجاع بتوفير هذه الخاصية لإنتاج الكثير من إمكانات البحث المتولدة عن استخدام التحليل الصرفي لكلمات اللغة العربية. تبدأ هذه العملية بإضافة المعاجم اللغوية الآلية لنظام الاسترجاع، ولا ينسحب التحليل الصرفي على كل كلمات اللغة

---

(١) هانئ محيى الدين عطية .**تكشيف الصور بين العلم والتعلم: ورقة في ملف القضية الفلسطينية**-. أسيوط: المؤتمر القومي الخامس لاختصاصيي المكتبات والمعلومات والأرشيف في مصر، ٢٠٠١م.

العربية وإنما يستثنى من ذلك الحروف وأدوات التركيب... وغيرها من الكلمات التي لا تدل بمفردها على معنى معين. ويعد الإدريسي /http://www.alidrisi.com وهو محرك بحث أنتجته شركة صخر للحاسبات من أهم نظم استرجاع المعلومات التي تعمل بتقنية التحليل الصرفي، مما أدى إلى بزوغ إمكانات بحث متقدمة لا ترى في غيره من محركات البحث العربية. لقد بات التحليل الصرفي الحل لمعالجة بحث اللغة العربية بعد تناوله من جانب الصوينع وغيره تعتبر محركات بحث الصور تربة خصبة تنمو بها تطبيقات التحليل الصرفي، حيث إن تحليل موضوعات الصور باستخدام اللغة الطبيعية يحقق الهدف المرجو في استرجاع الصور العامة.

وتتمثل تقنيات البحث المتقدمة المتولدة عن التحليل الصرفي للغة العربية فيما يلي:-

١/١/٢/٢/٣ البحث بالتطابق

يطلق على هذا النوع من البحث مسميات عدة منها البحث بالكلمة أو البحث بالمطابقة. وتعتمد هذه الطريقة على وضع كلمة محددة يبحث بها نظام الاسترجاع داخل محتوياته، على أن المضاهاة هنا اعتمادًا على نفس الشكل الذي أقره المستفيد لإجراء البحث؛ فإذا ما استخدم المستفيد مثلا كلمة (عربي) فإن النظام يعمل على استرجاع كل الوحدات التي تحتوي على كلمة عربي.. بينما لا يسترجع النظام وحدات الكلمات الأخرى مثل (عرب، عربة، عربية) . ويحتاج المستفيد إلى هذا النوع من البحث في حالة طلب معلومات عن موضوع له مخصص بحيث يهتم المستفيد بالتحقيق أكثر من الاستدعاء الذي ستكون نسبته في هذه الحالة قليلة جدًا.

٢/١/٢/٢/٣ البحث غير المطابق "البحث باللواصق"

يختلف هذا الشكل من البحث عن الشكل السابق في توسيع نطاق البحث داخل النظام. والبحث غير المطابق للكلمة مؤداه أن يسترجع النظام مختلف الأشكال للكلمة الواحدة، مع استدعاء أشكال الكلمة مع كل اللواصق التي تسبقها أو تليها في وحدات نظام الاسترجاع. فمثلا إذا أراد الباحث البحث عن كلمة (عرب)، فإنه سيحصل على (عرب، عربي، عربية، عربيات) وكذلك يسترجع جميع الكلمات التي تحمل سوابق مثل (العربي، وللعربي، وبالعربي) ومثلها للكلمات

الأخرى"[1] ويحتاج المستفيد لمثل هذا الشكل من البحث إذا ما أرا. الحصول على كل وحدات المعلومات حول موضوع محدد يتمثل في كلمة محددة.

٣/١/٢/٢/٣ البحث بجذر الكلمة "حل مشكلة الاشتقاق"

يأتي هذا الشكل في المستوى الثالث من توسيع نطاق البحث، وهو أهم أمثلة استخدام التحليل الصرفي للغة العربية، ويعني إجراء تحليل الكلمة إلى الحروف الأصلية بها واستخدامها نواة للبحث داخل نظام الاسترجاع. ومثالا على ذلك فإن باحثا عن معلومات تتعلق بـ (عرب) سوف تؤدي إلى استرجاع وحدات من داخل النظام تتضمن كلمات (عرب، عربي، عربية، عروبة،... إلخ) .[2] يؤدي هذا الشكل من البحث بطبيعة عمله إلى استدعاء عدد غير محدود من مصادر المعلومات، عملا على تجميع كل ما يتعلق بالجذر موضع البحث. ويترتب على نوع البحث نوع المستفيدين فالمستخدم لهذا الشكل من البحث يلجأ إليه عند الحاجة إلى كل ما يتعلق بالموضوع سواء من قريب أو بعيد نظرًا لما يمكن استدعاؤه من معلومات. ولكن يعيبه أنه قد يأتي بكلمات ليس لها علاقة بالموضع مثل (عراب، إعراب).

٤/١/٢/٢/٣ البحث بالتشكيل لكلمات المشترك اللفظي

يطلق على البحث بالتشكيل مسمى فك اللبس كما ورد في بعض مصادر الحديث عن أشكال البحث في اللغة العربية. ويخدم هذا الشكل المسفيدين في تفادي أحد أهم عوائق الاسترجاع في اللغة العربية وهو المشترك اللفظي للكلمات، مثل (عِلْم، علّم، علم) . وقد عانى المستفيدون من كثرة الاختلاط بين الأشكال المختلفة للكلمة الواحدة، وكان الحل البديل هو وضع كلمة أخرى بجنب الكلمة المختلطة للاسترشاد بمعنى الكلمة الأخرى. ويعطي هذا الشكل للمستيد أقصى درجات الدقة في الحصول على التحقيق المطلوب. وتتمثل طريقة استخدام شكل البحث في مجموعة بدائل لشكل الإعجام للكلمة يعرضها نظام الاسترجع من واقع التحليل الصرفي للغة داخل النظام.

---

(١) إرشادات البحث في الإدريسي. ٢٠٠٢/١٠/١٠، مسار http://asb.kacst.edu.sa/IDRADMIN/Arabic/Help/ATips.htm

(٢) مساعد بن صالح الطيار .**محركات البحث العربية وطرائق البحث فيها**، الجزيرة ٢٠٠١، تاريخ الاستشهاد ٢٠٠٢/١٠/١٠، مسار http:/www.al-jazirah.com.sa/evillage/22112001/wr1.htm

٥/١/٢/٢/٣ البحث بالمقاطع (أجزاء الكلمة)

يقوم هذا الشكل على إعطاء مجموعة من المعاملات أو الحروف التي يستخدمها الباحث في تشكيل إستراتيجية بحثه كما يريد، وتتركز هذه الحروف في علامات (؟، *) على أن استخدام هذه العلامات يختلف بين نظام وآخر، ويوجد جانبان من استخدام هذه الآلية في البحث هما استخدام معاملات البحث بالمقاطع في منتصف الكلمات، أو استخدام المعاملات في نهاية الكلمات؛ أما الجانب الأول من الاستخدام وهو في منتصف الكلمات فيأتي إما لعدم معرفة الباحث بالحروف الصحيحة المكونة لكلمات البحث، مثل البحث عن عرب تسترجع كلمات عراب، وأما الجانب الثاني من استخدام معاملات البحث الحر فيأتي في نهاية الكلمة ويستخدم للتخلي عن البحث بكلمات متعددة تشترك في أصل واحد مختلف الامتدادات وتستخدم العلامة (*) للتعبير عن مجموعة من الحروف التي تأتي في نهاية الكلمة مثل (عرب *) فإنها تسترجع " عربي، عربية، عربيات... الخ، مع التأكيد على أنها لن تسترجع كلمات مثل " العربي، وبالعربي... الخ [١].

٦/١/٢/٢/٣ البحث بحساسية الحالة

يعبر هذا الشكل عن المصطلح case sensitive ويطلق عليه في بحث الإدريسي مراعاة حالة الأحرف، ويستخدم المستفيد شكل البحث في حالة إذا ما أدت الأشكال المختلفة لحروف الكلمة معاني مختلفة، يعمل البحث بحساسية الحروف أو شكل الحروف على البحث عن اتجاهات موضوعية مختلفة لموضوع واحد ولكلمة واحدة ؛ حيث تعمل شكل الكلمة على بيان جانب من الموضوع مثل كلمة "أعلام" فإن استخدام البحث بدون حساسية الحروف يسترجع كلمات أخرى مثل "إعلام" . حيث يعامل حرف الألف كحرف واحد دون الأخذ في الاعتبار حالة الهمزة (ا، أ، إ، آ) وتستخدم هذه الآلية في محركات البحث متعددة الموضوعات، هذا إلى جانب اختيار شكل الحروف كواحدة من أدوات ضبط المصطلحات في عمليات التكشيف بما يصنع معه تحديدًا أكثر للمصطلحات الكشفية. ولم يتعامل أي من محركات البحث السابقة مع البحث بشكل الحروف، وهذا يرجع في رأي الباحث

---

(١) إرشادات البحث في الإدريسي. مصدر سابق.ص ٤.

إلى أن مثل هذه الآلية تحتاج إلى مرحلة متقدمة من البحث من جانب المستفيدين.. الأمر الذي لا يتناسب مع الاستخدام العام للويب.

٧/١/٢/٢/٣ البحث بالمترادفات

في هذا النوع من البحث يرتكز نظام الاسترجاع في إجراء البحث على المعجم الآلي، حيث تحديد المترادفات لكلمة البحث، واستخدامها في استدعاء مصادر المعلومات من وحدات النظام التي تحتوي على مرادفات كلمة البحث. فمثلا يؤدي البحث عن كلمة "تجهيز" إلى استرجاع جميع المستندات المفهرسة، التي تحتوي على كلمة (تجهيز) أو إحدى مترادفاتها مثل: إعداد، تهيئة... وغيرها من مرادفات الكلمة إن وجدت. ويمثل هذا النوع من البحث أهمية بالغة في البحث عن الصور حيث يمكن التعبير عن الصورة بالكثير من الكلمات. ويعد محرك البحث الإدريسي[1] من أفضل أدوات البحث العربية التي تعمل باستخدام تقنية التحليل الصرفي.بما يمكن معه إنتاج آليات بحث مثل البحث بالمترادفات.

٨/١/٢/٢/٣ البحث بالمتضادات

يستخدم المستفيدون هذا الشكل في البحث من أجل استرجاع مجموعة الوحدات التي تتعلق بالكلمات ومتضاداتها. وعلى ذلك، فإذا ما رغب المستفيد في البحث عن كلمة (السلام) ، ويختار البحث بالمتضادات، فإن نظام الاسترجاع يستدعي أيضًا كلمة (الحرب) وغيرها. وهذه الطريقة ربما تكون مفيدة في البحث في النصوص، حيث تأتي كلمات مرتبطة ببعضها مثل (الحرب والسلام) ، و(الأبيض والأسود) ، و(الخير والشر) و(الجنة والنار) ، ولكن بالنسبة للصور، فإن هذا ربما يستدعي صورًا مخالفة لما هو مطلوب. وأفضل أدوات البحث العربية في ذلك محرك البحث الإدريسي.

٩/١/٢/٢/٣ البحث بالمعاني

يوفر استخدام التحليل الصرفي البحث بالمعاني المتعددة التي توجد عليها الكلمة في اللغتين العربية والإنجليزية، ويعتقد الباحث أنه لا يوجد الكثير من المستفيدين من يرغب في هذا البحث إذا ما عرف معنى المصطلحات التي

---

(١) www.alidrisi.com

يستخدمها لبحث موضوع ما. وقد عثر الباحث عن هذا الشكل في خاصية البحث المتقدم لمحرك بحث عربي هو (عجيب www.ajeeb.com، ومع استخدام كلمة البحث (سيد) خرجت النتائج متضمنة كلمات تحمل معنى السيد أو السيادة بجانب استرجاع الوحدات التي تتضمن كلمة البحث ذاتها.

٣/٢/٢/١/١٠ التعرف إلى الأخطاء العربية الشائعة

يقصد بهذا الشكل من البحث التعرف إلى الأخطاء التي توجد عليها كلمات اللغة العربية سواء في بناء النص واسترجاعه. وللغة العربية طبيعة تختلف عن تلك اللغات الغروية أو اللصقية مثل اللغات اللاتينية والإنجليزية. وتعد الهاء الختامية والتاء المربوطة والياء المقصورة وهمزة الوصل والقطع من أكثر الأخطاء شيوعًا بين الكتاب والمستفيدين، ومثال ذلك فإن كلمة (احمد) تسترجع معها الشكل الآخر من الهمزة وهو (أحمد). ويجب في هذه الحالة تعطيل خاصية البحث بحساسية الأحرف؛ لأنها تعمل بشكل عكسي لهذا النوع من البحث.

٣/٢/٢/٢ آليات البحث بكلمات متعددة

عند استخدام البحث بالكلمات المتعددة، يستقي المستفيد من إدراكه بالموضوع مجموعة من الكلمات التي تتكامل معًا لأجل التعبير عن الموضوع، وكلما زاد عدد الكلمات كان ذلك أحرى بتضييق نطاق البحث ومن ثم زيادة التحقيق. وآليات البحث المستخدمة مع الكلمات المتعددة هي :-

٣/٢/٢/٢/١ البحث في الحقول

يستخدم المستفيد في هذا الشكل من البحث مجموعة من الحقول المعدة سلفًا في محرك البحث وعادة تكون هذه الحقول محدودة أو بمعنى آخر شائعة الاستخدام كأن يبحث المستفيد في حقل المؤلف أو العنوان أو الناشر أو السنة، وفي التسجيلة الرقمية للصورة يمكن البحث عن صورة محددة بتكامل حقولها معًا في البحث. وتستند إستراتيجيات البحث بحقول التسجيلة إلى دعامة المنطق البوليني في تكوين العلاقات بين الكلمات في الحقل الواحد وبين الحقول المختلفة وبعضها الآخر. وتحدد المعاملات العلاقات التالية؛ فعند استخدام معامل البحث AND يرمز ذلك إلى الجمع بين الكلمتين أو الحقلين في الوحدات المسترجعة، مما يكون له أثر تضييق نطاق البحث، أما معامل OR فيرمز إلى الحصول على أي من

الحقلين أو الكلمتين داخل الوحدات المسترجعة، مما ينتج عنه توسيع نطاق البحث.
أما المعامل الثالث هو NOT فيرمز إلى إبعاد الكلمة أو الحقل من بين الوحدات التي يتم استرجاعها.

٢/٢/٢/٢/٣ صياغة الجملة البحثية

تتيح بعض محركات البحث للمستفيد حرية صيغة الجملة البحثية بشكل أكثر مرونة من تلك الموجودة في الحقول حتى يصل بها إلى الصورة المطلوبة، وبالتالي، فإن صورة ملونة عن القدس التقطها فنان محدد في وقت محدد يجعل لهذه الصورة الأهمية الكبيرة، يسعى المستفيد إلى صياغة إستراتيجية بحث من حقول ثابتة يصل بها إلى الصورة المطلوبة، وتكون هذه الإستراتيجية على شكل SU=() or SU=() and Au=() and CO=(color) and DT=(1948). حيث su الموضوع، au الفنان، dt التاريخ، co اللون. ويمكن في هذا النوع من البحث استخدام معاملات إضافية غير معاملات المنطق البوليني مثل معاملات حالة المجاورة بين كلمات البحث، والبحث بالتقارب في إحدى حالاته يستخدم عنصر التطابق بين كلمات البحث المدخلة والنتائج المسترجعة ؛ حيث تحتوي النتائج على نفس الكلمات والترتيب التي تم البحث بها. لكن الاختلاف بين هذين الشكلين يأتي في الحالات الأخرى من مدى التقارب المطلوب بين الكلمات، ويستخدم نظام الاسترجاع معاملات للتحكم في البحث مثل (near *) ويعني قريبًا بدرجة محددة وفقا لما يكون عليها عدد العلامات *، أما "adjacent, allowed by" يعني بجوار أو ملتصقًا ويستخدم في ربط كلمات البحث معا. ويؤثر البحث بالتقارب في ترتيب النتائج النهائية وليس فقط استدعائها، فكلما كانت كلمات البحث متقاربة اقتربت وحدات المعلومات إلى الأرقام الأولى لنتائج البحث.

٣/٢/٢/٢/٣ استخدام علامات التطابق

يستخدم هذا النوع من البحث للتأكيد على استخدام الكلمات المطلوب البحث عنها بنفس الترتيب المدخلة به. وتستخدم محركات البحث علامتي التنصيص " " ليلزم نظام الاسترجاع باستدعاء النتائج مشتملة على كلمات البحث بنفس الترتيب المدخلة به. مما يعني ضمنًا اختيار المستفيد لخيار التطابق بين كلمات البحث المدخلة والنتائج المسترجعة من النظام.

٤/٢/٢/٢/٣ البحث المعقد أو المركب

لا يعد البحث المركب أو المعقد أحد أشكال المعالجة في اللغة العربية، إنما يعد هذا الشكل استخدامًا لمجموعة من أشكال البحث السابقة في صياغة إستراتيجية واحدة للبحث داخل نظام الاسترجاع. ويأتي استخدام هذا الشكل في بحث الموضوعات التي تحتوي على جوانب موضوعية مختلفة. مثال ذلك كأن يكون البحث عن مصادر تتوفر فيها الشروط التالية مجتمعة، وهي أن تحتوي على أي من الكلمات: (دولة، جمهورية، مملكة، سلطة، حكومة) بمترادفاتها، وأن تحتوي على أي من الكلمات: (صحافة، إعلام، رأي) بلواصقها، إلى جانب أن تحتوي على كل الكلمات: (عرب، سجن) بمشتقاتها، وفي ذات البحث لا تحتوي على كلمة (إسرائيل) بلواصقها. وبتطبيق النظام لهذه الشروط معًا فإنها تؤدي إلى الحصول على مصادر خاصة بحرية الرأي والصحافة في البلدان العربية.

ويتكامل عُنْصُرا الوصف المادي والوصف الموضوعي مع آليات بحث النص العربي في تكوين نظام استرجاع للصور باللغة العربية، من شأنه إفادة الباحثين عن أوعية الصور للأغراض المعلوماتية المختلفة. وتأتي مقاييس استدعاء التحقيق والشوشرة ضوابط تحسن من أداء نظم استرجاع الصور. يتعرف المستفيد ومصمم النظام على مدى الكفاية التي يعمل بها محرك أو قاعدة بيانات الصور. والفصول التالية تعرض لعينة من محركات الصور وخصائص الإضافة والتكشيف والبحث داخلها، كاشفة عن العناصر الأخرى المكملة لأداء محرك البحث من واجهات وعرض النتائج... الخ.

٣/٢/٢/٣ أساليب بحث أخرى متطورة

طورت نظم الاسترجاع من أساليب البحث التقليدية المستخدمة في معظم محركات البحث وذلك من خلال استخدام أساليب أخرى أكثر مرونة من أهمها:

١/٣/٢/٢/٣ البحث بصيغة توسيع السؤال

يتم هذا الشكل من البحث بأن يقوم الباحث بكتابة الكلمة أو الكلمات التي يرغب فيها ثم بعد الحصول على النتائج يتاح للباحث الاختيار من بين نتائج البحث الأكثر مطابقة لاحتياجاته، ثم يقوم بإجراء البحث مرة أخرى، وعندها يقوم محرك البحث بإضافة الكلمات الموجودة في نتائج البحث المختارة والبحث بها لاستدعاء

نتائج أخرى أكثر دقة. ويفيد هذا النوع من البحث في محركات بحث الصور بدرجة كبيرة، حيث إن البحث قد يأتي بصور ليس لها أي علاقة برغبة المسفيد ولكن عند اختيار مجموعة من الصور وإجراء البحث مرة أخرى يمكن أن تضاف الكلمات المفتاحية المصاحبة لها إلى الجملة البحثية واستدعاء صور أكثر مطابقة لرغبة الباحث.

١/٣/٢/٢/٣ البحث بوزن الكلمة

قد تنصرف نظم الاسترجاع إلى احتساب تردد الكلمات لحساب الأهمية التي يكون عليها مصدر المعلومات في تلبية المعلومات موضع البحث، أيضًا فإن نظم الاسترجاع تحتكم إلى مكان الكلمات المفتاحية في مصادر المعلومات سواء في العنوان أو الرموز المعيارية... الخ؛ إلا أن البحث بتقدير الأهمية يعطي المستفيد إمكانية الاسترجاع بحسب أهمية الكلمات المفتاحية التي استخدمها للبحث(١). وهذا النوع من البحث أكثر مناسبة مع الكلمات المصاحبة للنص منها لتلك المصاحبة للكلمات المفتاحية.

تعتمد فكرة البحث بوزن الكلمة على مطابقة الكلمات الواردة في عنوان الوثيقة أو النص مع الكلمات الواردة في السؤال و إعطاء كل كلمة منها وزنًا بحسب أهميتها و ذلك وفق الخطوات التالية :

١. يقارن النظام بين الكلمات الواردة في السؤال و الكلمات الواردة في النص أو العنوان.

٢. تستبعد الوثائق التي لا تحتوي على أي عنصر من عناصر السؤال.

٣. ترتب الوثائق التي تحتوي على عنصر واحد أو عنصرين أو أكثر من عناصر السؤال حسب وزن الكلمات الدالة فيها و يتم حساب وزن الكلمات بإحدى الطريقتين:-

❖ في النظم البسيطة تعطى الكلمة قيمة (١) إذا كانت تحتوي على عنصر ورد في السؤال و عنوان أو نص الوثيقة.و قيمة (صر) إذا لم تكن في العنوان أو النص و بالتالي يحسب النظام وزن الوثيقة

---

(١) إرشادات البحث في الإدريسي. مصدر سابق. ص ١٠.

الإجمالي بالنسبة لطلب معين بعدد الكلمات المتوفرة بها من مجموع الكلمات الدالة المطلوبة في السؤال و ترتيب الكلمات الأكثر ملاءمة أو مضاهاة إلى الأقل ملاءمة أو مضاهاة.

- ❖ في النظم المتطورة يقوم الباحث بإعطاء وزن لكل كلمة يعبر من خلالها عن أهميتها بالنسبة له و يحسب وزن الوثيقة و مدى ملاءمتها للطلب بالنسبة إلى عدد الكلمات المتوفرة وأوزانها المختلفة في السؤال ويحسب الوزن كما يلي :-

وزن الكلمة أ في الوثيقة x وزن الكلمة أ في السؤال + وزن الكلمة ب في الوثيقة x وزن الكلمة ب في السؤال.

٤. تحسب أوزان الكلمات الواردة في السؤال آليًا متناسبة عكسيًا مع معدل تواتر الكلمة في مجموعة النصوص في تخصص واحد أي أن الكلمة نادرة الاستخدام في مجموعة النصوص يفترض أنه تعبر بدقة أكبر عن حاجة الباحث من تلك الشائعة في الاستخدام.

٣/٣ الخلاصة

تتسم الصور الرقمية بمجموعة متنوعة من ملامح الوصف التي تختلف بين وصف الشكل إلى وصف محتوى الصور بجانب ملامح وصف ملف الصور في ذاته مثل حجم ونوع ملف الصورة. وقد تبنت الدراسة معيار دبلن كور للمعالجة الفنية للصور الرقمية على الويب. غير أن المعالجة الفنية للصور الرقمية لا تتوقف فقط عند الوصف المادي، و إنما تتعدى إلى تحليل محتوى الصورة بعناصره المختلفة، وقد اعتمدت الدراسة على تبني نموذجين من الصورة لتحليلهما بما يوضح ملامح تكشيف الصور الرقمية على الويب. وتأتي اللغة العربية لتمثل البيئة التي يعمل بها محرك البحث ؛ حيث يتم التعرف إلى طبيعة ما تحمله اللغة العربية من خصائص، إلى التعرف إلى التقنيات و آليات البحث التي تحل بها اللغة العربية مشكلاتها في البحث على الويب عامة و بحث الصور خاصة.

الفصل الرابع

محركات بحث الصور

٤/٠ التمهيد

٤/١ آليات البحث في محركات الصور

٤/٢ خصائص محركات بحث الصور

٤/٣ خدمات محركات بحث الصور

٤/٤ كفاية استرجاع محركات بحث الصور

٤/٥ الخلاصة



محركات
بحـــث الصـــــور

٤/٠ التمهيد

تشترك كل من محركات بحث الويب والمحركات محددة البحث في السمات العامة لعمل محرك البحث في شكله التقليدي وذلك فيما يتعلق بالبحث الخارجي للمصادر، التنظيم والاسترجاع: إلا أن هناك الكثير من خصائص العمل التي تختلف بين طبيعة بحث فضاء الويب وبين الملفات الرقمية التي تعتمد على تحديد آليات البحث المسبق؛ مثل الاعتماد على بعض حقول تخزين واسترجاع المعلومات أو البحث في حقول Meta Tags. وسوف يتناول هذا الفصل آليات العمل في محركات البحث وخصائصه، وخدماته وكفاية الاسترجاع به.

٤/١ آليات البحث في محركات الصور

يعتمد هذا القسم على التفريق النوعي بين محركات بحث العينة جميعها وبين محركي البحث Corbis, Freefoto؛ حيث إنهما يمثلان نوعًا خاصا من محركات بحث الصور يقترب في درجة عمله إلى درجة قواعد البيانات، هو ما نتج عن استخدام العنصر البشري في هذين المحركين.

٤/١/١ آليات البحث العامة

تعتمد محركات البحث بصفة عامة ومحركات بحث الصور بصفة خاصة على استخدام نوعين من أشكال استرجاع محتوياتها من مصادر المعلومات، يستخدم أول الأسلوبين في عرض محتويات قاعدة بيانات محرك البحث أو نظام الاسترجاع بالاعتماد على وضع تقسيم لموضوعات قاعدة

البيانات اعتمادًا على استخدام المصطلحات ورؤوس الموضوعات كما في حالة قواعد البيانات النصية، أو استخدام تمثيل مصغر لمصادر المعلومات نفسها وخاصة المعلومات المرئية كما في حالة عرض لقطات صغيرة من الصور. على هذا الأسلوب اسم التصفح أو التقسيم الموضوعي والذي يعتمد في أغلب الأحيان على خطة تصنيف لموضوعات المعرفة مثل خطة تصنيف ديوي العشري أو خطة التصنيف العشري العالمي. أما الأسلوب الآخر في البحث فهو الاعتماد على البحث الحر المتمثل في استخدام الكلمات المفتاحية من جانب المستفيد دون التقيد بمصطلحات محددة يعرضها نظام الاسترجاع كما في تقسيم المصطلحات السابقة، ويرتبط بهذا الجانب استخدام تقنيات وآليات بحث تلائم الصيغ المختلفة التي يدخل بها المستفيد استفساره إلى قاعدة البيانات، خاصة تلك التي تعمل بلغات عالية الدقة مثل اللغة العربية. إن التقسيم بين أساليب البحث في أدوات بحث الويب هو الأصل الذي يبنى عليه الفارق بين محركات البحث وأدلة البحث، فان محركات البحث تعتمد على التكشيف الكامل لصفحات الويب، مما يسهل البحث الحر بالكلمات المفتاحية بخلاف ما تكون عليه أدلة البحث من الاعتماد على التكشيف الإنساني الذي يصف الموضوع في مصطلحات محددة يلزم المستفيد بالبحث بها من خلال قوائم الموضوعات مختلة التخصص والمتفرعة بعضها من بعض.

ويتعرض الفصل هنا إلى استخدام كلا الأسلوبين في محركات بحث الصور والكفاية التي يمكن أن يعمل بها كلا النوعين. ولم تعتمد مجموعة محركات بحث الويب على استخدام أسلوب التصفح في تنظيم واسترجاع ملفات الصور الرقمية، وإنما يرجع ذلك إلى الطبيعة العامة لمحرك البحث التي يتبناها في تنظيم واسترجاع جميع أشكال مصادر المعلومات على الويب، على الرغم من أن الكثير من محركات بحث الويب العامة قد لجأت إلى استخدام صيغة دليل البحث للجمع بين الحسنيين؛ أسلوب التصفح الرئيس من ناحية وأسلوب التصفح من ناحية أخرى وعلى رأسها Google. ويوضح الجدول رقم (٤-١) النمط المتبع في كل محرك من عينة الدراسة :-

الجدول رقم (٤-١) يوضح أساليب استرجاع المعلومات لمحركات بحث الويب

محركات البحث		محركات بحث الويب						محركات بحث محددة	
		Altavista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Freefoto
أسلوب التصفح	قائمة واحدة	--	--	--	--	--	--	مصطلحات عامة	--
	قوائم متعددة	--	قائمتين	--	--	--	--	--	ثلاث قوائم
أسلوب الكلمات المفتاحية		مزود بآليات بحث	يقتصر على وضع الكلمات	مزود بآليات بحث	مزود بآليات بحث	مزود بآليات بحث	مزود بآليات بحث	مزود بآليات بحث	يقتصر على وضع الكلمات

ويوضح الجدول رقم (٤-١) الأساليب المتبعة في استرجاع الصور داخل محركات بحث الصور، وترتكز محركات بحث الصور على الكلمات المفتاحية كأسلوب أساس للبحث، على حين يتضح الإهمال شبه التام لأسلوب التصفح، وقد شذ عن ذلك فقط محرك البحث Ditto بالاعتماد على مستويين من الموضوعات العامة التي تغطي مجموعات الصور داخله؛ إضافة إلى أن محرك البحث Google لم يمتد دليله البسيط إلى التعامل مع ملفات الصور حيث التركيز على تنظيم صفحات الويب فقط. وتعتمد محركات البحث على استخدام الكلمات المفتاحية التي تتناسب مع طبيعة تكوين قواعد تلك المحركات من خلال التقاط الصفحات بلغتها الطبيعية وتحديد نقاط ومداخل استرجاعها بتكشيف النص الكامل لمواقع صفحات الويب. وعلى ذلك يمكن أن نخلص إلى أن الفارق بين تلك المحركات يأتي فقط في آليات البحث الحر بالكلمات المفتاحية وكيفية صياغة إستراتيجيات البحث لبحث قواعد بياناتها، ويأتي ذلك في الحديث عن آليات البحث داخل محركات الدراسة.

كما يتضح من الجدول رقم (٤-١) أن كلا من محركي البحث قد اعتمد على الجمع بين أسلوبي البحث؛ وإنما يرجع ذلك في رأي الباحث إلى التدخل البشري لتنظيم ملفات الصور، وتوقع المستوى الأدنى من مهارات مستخدمي الويب في التفاعل مع البحث بالكلمات المفتاحية بمفردها. غير أن كلا المحركين لم يغفل أسلوب البحث الحر كأسلوب أساس سريع في إجراء البحث عن ملفات الصور داخل جميع أنواع محركات البحث. ومن ناحية أخرى فقد جاءت قوائم الموضوعات المستخدمة في محركي Freefoto, Corbis على غير الدقة المطلوبة أو الحصر الذي

يتماشى مع تعدد موضوعات الصور العامة، احتوى أحدهما على قائمة واحدة تضم الموضوعات العامة للبحث غير مرتبطة فيما بينها بأي أساس للتقسيم، واحتوى الآخر على ثلاثة قوائم تضم الأساسية الموضوعات العامة. أما القوائم الفرعية فتضم الموضوعات الفرعية للقائمة الأساسية.

١/١/١/٤ آليات البحث بالتصفح

إن استخدام التصفح بالمصطلحات أو الكلمات المفتاحية يعتمد في أول الأمر على حجم قاعدة بيانات الصور، فإذا ما كانت قاعدة البيانات تضم مجموعات كبيرة من الصورة فإنه من الجدير استخدام إحدى الأدوات الحصرية لرؤوس الموضوعات سواء أكانت الكشافات الهجائية لخطط التصنيف أم قوائم رؤوس الموضوعات، حيث يعمل ذلك على تضييق الموضوعات بدرجة كبيرة أمام الباحثين عن المعلومات المتخصصة والصور المرتبطة بموضوع محدد، مما يوفر على المستفيد معاركة كلمات البحث المفتاحية والشكل الملائم منها للبحث. والتصفح باستخدام المصطلحات يختلف بدوره عن التصفح باستخدام الصور المصغرة، والذي تناسب بطبيعته مع قواعد بيانات الصور ذات الحجم الصغير. ومن جانب آخر فإن التصفح باللقطات أكثر ملاءمة للتعرف إلى الأوعية المرئية مثل الصور الثابتة، حيث لا يكون النص بمصطلحات أو نص وسيط بين الصورة والمستفيد، مما يمكن الصورة بذاتها من أن تعكس مختلف الموضوعات التي تحتويها. وختامًا فإن أسلوب التصفح يعد أسلوبًا مساعدًا لمحرك البحث يمكن الاستفادة منه في تدعيم البحث بالكلمات المفتاحية الذي يعد أساس عمل محركات البحث ليس في الاسترجاع وإنما في التكشيف للغة النصوص الطبيعية.

إن ناتج دراسة أساليب الاسترجاع يفيد أن الصور الرقمية الطبيعية تحتاج إلى طبيعة تقسيم تختلف عن مثيلاتها في الصور الإعلامية والصور العلمية؛ فالصور العامة تحتاج إلى الاعتماد على خطة تصنيف حصرية يمكن أن تضم مجموعات الصور في مختلف الموضوعات، أو أن تكون تلك الأداة هي إحدى قوائم رؤوس الموضوعات التي تمكن المستفيدين من متابعة الترتيب الهجائي ومتابعة رؤوس الموضوعات المقننة التي تملك الروابط إلى الصور المعبرة عنها في المحتوى، غير أنه من المسلم به أن ترتبط الصورة العامة الواحدة بكل رؤوس

الموضوعات التي تمثلها نظرًا لطبيعة الصور العامة. كما جاء في الجدول رقم (٤-١) من توضيح لأساليب البحث الأساسية في محركات الصور المحددة.

٢/١/١/٤ آليات البحث بالكلمات المفتاحية

تستخدم محركات البحث مجموعة من الآليات التي تساعد المستفيدين في صياغة شكل إستراتيجيات البحث عن مصادر معلومات الويب النصية والمرئية، واعتمادًا على أن الكلمات المفتاحية هي كلمات ومصطلحات لفظية، فإن تأثرها باللغة التي يبحث بها محرك البحث تنعكس على مجموعة الآليات التي يتم استخدامها لتناسب عملية بحث النصوص. وتختلف آليات بحث كلمات ومصطلحات اللغة العربية عن تلك المستخدمة في بحث المصطلحات الإنجليزية، وهو ما يتم دراسته اعتمادًا على محركي البحث Google, Fast لتبنيهما اللغة العربية في بحث الصور.

وتعطي آليات البحث إمكانية التحكم في صنع أشكال مختلفة من البحث بمجموعة واحدة من الكلمات أو كلمات مفردة. وتختلف إستراتيجيات البحث بدورها وطبيعتها عن آليات البحث التي تتكامل لتحقيق أنواع أشكال مختلفة من الإستراتيجيات. وتتعدد إستراتيجيات البحث المستخدمة لبحث محركات البحث بين البحث الحر بالكلمات المفتاحية واستخدام الحقول في بحث مصادر وصفحات الويب، وبنظرة إلى أعلى قليلا فإن الويب يتم بحثها من خلال مجموعة مختلفة من إستراتيجيات البحث مثل استخدام البحث بالمسار، أو استخدام الأدلة الموضوعية،أو استخدام محركات وكشافات البحث على اختلاف أنواع هذه المحركات بين محركات البحث المتخصصة ومحركات البحث العامة أو المتعددة التي تبحث في محركات أخرى Meta search Engines أو استخدام قواعد البيانات على الويب والتي تعرف الويب الخفية Invisible Web الخدمات التقيميية للمعلومات على الويب ومساعدات الويب في الاستشهادات المرجعية. وقد تم حصر مجموعة آليات البحث المختلفة مع اختبار تواجد وكفاية تطبيق هذه الآليات في محركات الدراسة.[١] وتمثل آليات البحث التالية كل ما تعتمد عليه محركات البحث على

(١) Lazewski, Barbara. WWW Search Strategies, University of Wisconsin-Madison, 2002, visited 30/11/2002, (http://www.library.wisc.edu/libraries/Steenbock/services/wwwstrat.htm)

الويب سواء منها التي تبحث في النصوص أو تلك التي تبحث في الوسائط الأخرى، كما صدق على ذلك[١] وجاءت كما يلي :-

يعتمد الجدول على قياس آليات البحث ثم توضح علامة وجود آلية البحث كاملة نشطة في المحرك، أما علامة x فتعني عدم وجود آلية البحث على الإطلاق.

الجدول رقم (٤-٢) يوضح آليات البحث المستخدمة في عينة محركات البحث

آليات البحث	محركات بحث الويب						محركات بحث محددة	
	Alta- Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
البحث بالتطابق	✓	x	✓	✓	x	x	✓	✓
البحث الحر (?,*)	✓	x	x	x	x	x	✓	✓
البحث حساسية الحروف	x	x	x	x	x	x	x	x
المنطق البوليني	✓	✓	x	✓	✓	x	✓	x
البحث بالعبارات والتوقف	✓	✓	✓	✓	x	x	✓	x
البحث ببدائل المنطق البوليني (+ و-)	✓	✓	x	x	✓	✓	✓	x
البحث بالتقارب	x	x	x	x	x	x	x	x

ويعطي الجدول رقم (٤-٢) المحتوي على محركات بحث الصور وآليات البحث المستخدمة مدلولات يمكن توضيحها فيما يلي :-

- استخدم ثلاثة محركات للصور آلية البحث بتطابق الكلمات وهي محركات Altavista, Google, Fast. ودلت النتائج المسترجعة داخل هذه المحركات على استخدام التطابق بأن كانت النتائج في أقل درجاتها وهو ما ينتج عن شدة التضييق الناتجة عن استرجاع عدد من الكلمات بعدد وترتيب واحد، وعلى الجانب الآخر، فإن استخدام علامات التطابق في المحركات الأخرى تساوى في

(١) ❖ تناولت آليات البحث على الويب مجموعة من الدراسات؛ إلا أن الباحث قد لاحظ التركيز بها على مجموعة واحدة من آليات بحث النصوص. ومن المصادر.

النتائج المسترجعة مع البحث باستخدام معامل البحث And إن وجد، أو الربط الذاتي داخل محرك البحث، وهو ما دل على غياب فاعلية معامل التطابق.

استخدم محرك Altavista معاملات البحث الحر. ويستخدم هذا المحرك معامل البحث * للبحث للتعويض عن الحروف المفقودة في امتداد كلمات البحث. وتعتبر معاملات البحث الحر من آليات البحث القوية في محركات البحث إذا ما تم توضيحها إلى المستفيدين، ولعل أكثرها إفادة استخدام معامل البحث * في نهاية كلمات البحث. ويفضل الباحث استخدام معاملات البحث الحر بشكل ثابت في وظائف الاستخدام. ويمكن للمستفيدين البحث الحر في التغلب على الأشكال المختلفة من الكلمات التي تشتق من أصل واحد، وهو أكثر ما يستخدم في اللغات اللصقية مثل اللغات اللاتينية بأنواعها، كما يستخدم هذا النوع من البحث إذا ما أراد المستفيدون الابتعاد عن المشكلات الهجائية التي تعتري اللغات الأخرى. كما يوضح الجدول رقم (٤-٢) استخدام محرك Freefoto لمعامل * في نهايات الكلمات وداخل كلمات البحث في حين أن معامل ؟ يستخدم فقط في داخل كلمات البحث، أما محرك البحث Corbis فانه يستخدم معامل البحث * فقط داخل حروف الكلمات. ومن الملاحظ من مقارنة نوعي محركات البحث أن محركات البحث المحددة تعتمد أسلوبًا خاصًا لآليات البحث خلاف آليات البحث عامة الاستخدام التي تعتمدها محركات بحث الويب حيث تتلاءم أكثر مع طبيعة مستخدميها.

لا توجد محركات بحث للصور تدعم آلية البحث بحساسية الحروف؛ وذلك يرجع في رأي الباحث إلى طبيعة محركات البحث العامة في موضوعاتها مما لا يجعلها تتوقف عند الجوانب المتخصصة للموضوع الواحد، والذي يمكن أن تعكسه حالة حروف المصطلح (الحروف الكبيرة EARTH أو الحروف الصغيرة earth). هذا فضلا عن أن هذه الآلية تحتاج إلى التجهيز الإنساني للمعلومات داخل محركات البحث، وهو ما يخالف طبيعة محركات البحث في اعتماد منهج وسياسة التكشيف الآلي بالاشتقاق لكامل متن صفحة الويب، مما يعني أن وجود الأشكال المختلفة للكلمة الواحدة يتوقف على مصمم وكاتب نص الصفحة المكشفة. فإذا عمد كاتب صفحة الويب إلى كتابة الكلمة بالحروف الكبيرة، فإن ذلك يعني الإشارة إلى جانب محدد من جوانب الموضوع لا يتم

استرجاع الوثيقة به إلا بالتعبير عن هذا الموضوع بالشكل السابق في الوثيقة. وقد لاحظ الباحث استخدام بعض محركات البحث مثل Altavista للبحث بآلية شكل الحروف فقط عند بحث صفحات الويب، في حين أنه لا يتم ذلك مع باقي الأشكال سواء أكانت ملفات الصوت أم ملفات الصور الرقمية. وذلك إنما يعكس في رأي الباحث تبني محركات البحث لمبدأ البحث بالشكل البسيط لملفات الصور والمتناسب مع إمكانات المستفيدين المتواضعة في البحث، بجانب افتقار الصور إلى التقنين والوصف الذي ينتج عنه آليات بحث متقدمة للاستدعاء.

- استخدمت أربعة محركات هي Picsearch, Google, Ditto, Altavista من مجموع ثمانية محركات، آلية المنطق البوليني في بحث الكلمات المفتاحية داخل قاعدة البيانات. دل هذا العدد في بداية الأمر على جنوح محركات بحث الصور عن استخدام المنطق البوليني لما يشوبه من نقص ممارسة وكفاية لمستخدمي الويب ومحركات البحث خاصة، وقد عملت بعض المحركات مثل Google على استخدام معاملات المنطق البوليني بما يتناسب مع مهارات المستفيدين المتفاوتة؛ حيث استخدم Google كلمة (كل) مقابلًا لمعامل البحث And، واستخدم كلمة "أي من" مقابلًا لمعامل البحث Or، واستخدم كلمة " ماعدا " مقابلًا لمعامل البحث Not. إلى جانب قيام بعض المحركات بالتعامل مع مبدأ الربط بين كلمات البحث جميعًا بشكل ذاتي بوضع معاملات تقابل معامل البحث And. وقد أوضح اختبار معاملات المنطق البوليني في محركات بحث الصور أنه مازالت الثقة بين محركات البحث والمستخدمين لا ترقى إلى الإفصاح عن معاملات البحث المتقدمة، كما أن التغلب على ذلك يكمن في استخدام مثل هذه المعاملات بالشكل الذي يتقبله المستوى الأقل من المستخدمين كما في حالة محرك Google. ويمكن تقديم المساعدة والتوضيح كعامل مساعد في فهم هذه المعاملات. إن استخدام معاملات المنطق البوليني من أهم مظاهر القوة في البحث داخل محركات البحث مما لا يتسم بغياب مثل تلك المعاملات في بحث مصادر المعلومات، وخاصة أوعية الصور الثابتة التي تحتاج في أغلب الأحيان إلى نص شارح لمختلف موضوعات الصورة الفوتوغرافية. ولقد سارع محرك البحث Corbis إلى اللحاق بتقدم

عمليات البحث على الويب، وقد اعتمد Corbis على إضافة آلية لبحث بالمنطق البوليني التي لم تكن إحدى آليات بحثه، حيث تم تفعيل معاملاته المختلفة and, or, not بمفاهيمها. وذلك يعكس الثقة الآخذة في التزايد بين قدرة محركات بحث الويب وبين مهارات مستخدميها.

توقف محركا Picsearch, Visoo عن العمل بآلية البحث بالعبارات وهو ما ظهر في اختلاف النتائج بين طرح كلمات البحث خالية من بعض الحروف، وطرح الكلمات نفسها في سياق سؤال للبحث. على الجانب الآخر فقد عملت محركات Altavista Google, Fast, Ditto، ظهر ذلك باختبار أعداد النتائج التي تساوت بين طرح كلمات البحث مفردة وبين طرحها في سؤال بحث يحمل داخله مجموعة من أدوات الربط بين الكلمات. وتعامل محرك Corbis فقط مع البحث بالعبارات حيث أهمل محرك البحث الحروف وكلمات الربط بين كلمات البحث ولم يؤثر ذلك على أعداد نتائج البحث.

استخدمت بعض محركات البحث معاملات المنطق البوليني على الشكل غير المعهود لتلك المعاملات، وقد فضل الباحث الحديث عن تلك المعاملات نظرًا لشيوع استخدامها بين محركات بحث الصور خاصة ومحركات البحث عامة. ويرجع الباحث استخدام محركات البحث لهذه الآليات إلى الثقة غير العادية في قدرة مستخدمي محركات البحث على التعامل الصريح مع معاملات المنطق البوليني. وتتركز بدائل المعاملات البولينية في (+) حيث يعمل المعامل الأول (+) على أداء مهمة And في الربط بين كلمات البحث الواحد لاسترجاعها مجتمعة في وحدة معلومات واحدة، أما معامل (-) فإنه يعمل على أداء مهمة Not في إبعاد ما يليها من كلمات مفتاحيه والحصول على وحدات معلومات لا تحتوي على هذه الكلمات. وإذا ما اعتمدت محركات البحث على معاملات المنطق البوليني الأساسية، فلا حاجة لها لاستخدام مثل هذه الآليات. ويتضح من الجدول رقم (٤-٢) لآليات البحث أن اثنين من هذه المحركات Picsearch, Ditto يعتمدان على تدعيم المنطق البوليني بمثل هذه المعاملات كي يتيحا للمستفيدين الصياغة البسيطة لإستراتيجيات البحث، في حين أن محرك البحث Altavista لا يعتمد على المنطق البوليني في البحث، إضافة إلى استخدامه للمعامل البديل (+) للربط بين كلمات البحث، ويعكس

ذلك لجوء مثل هذه المحركات إلى أبسط درجات البحث وأقلها كفاية في الربط بين مختلف الكلمات الداخلة لفراغ البحث دون النظر إلى الآليات الأخرى التي ترفع من كفاية الاستدعاء والتحقيق للنتائج المسترجعة. ولا تعتبر مثل هذه المعاملات بالمهمة إذا ما استمر التعامل المتزايد ولمتقدم مع محركات البحث، بل أن الأمر يستدعي التحرك السريع من لدن محركات البحث لتحديث وإيجاد آليات بحث أكثر كفاية، مما هي عليا الآن. ويعد استخدام معامل الربط + من جانب Corbis هو البديل التقني لوحيد لدى محرك البحث لاسترجاع نتائج تتعلق بموضوع مكون من أكثر من كلمة، إذا ما لم يتم اعتماد آلية المنطق البوليني في بحث مجموعاته من الصور الرقمية.

- لم تستخدم أي من محركات بحث العينة السابقة معاملات البحث بالتقارب. ويرجع الباحث هذا إلى الدرجة المتقدمة من البحث التي تعمل بها هذه الآلية والذي لا يتناسب حتى الآن مع مستخدمي محركات بحث الصور، التي مازالت برمتها تعاني من النقص في مختلف جوانبها سواء كان البناء أو التنظيم أو عملية الاسترجاع مرة أخرى. ويرى الباحث أن أهمية معاملات البحث بالتقريب تزداد يوما بعد الآخر؛ ذلك لما نراه من إمكانات فائقة ومتزايدة في بحث النص الكامل لصفحات الويب وهو ما تعمل به برامج التكشيف في محركات البحث، وتزداد فاعلية البحث بالتقريب أكثر من معاملات And و+ للربط في أن المعاملات السابقة تستدعي الوثائق التي تحتوي على كلمات البحث بدون الحكم إلى مدى التقارب بين المصطلحات الذي يتسع باتساع النص الكامل لصفحة الويب.

وقد أظهر الملحق رقم ٦ من ملاحق الدراسة نتائج البحث من وحدات الصور حسب آليات البحث المستخدمة.

٢/١/٤ آليات البحث النوعية

وهي مجموعة الآليات التي توجد بشكل منفرد في أي من محركات بحث الصور، وتتمثل فيما يلي:-

١/٢/١/٤ البحث بالتغطية الزمنية والمكانية لوحدات الصور

يهدف الباحث من عرض التغطية الزمنية والمكانية إلى إشراك سنوات

وأماكن نشر الصور الرقمية على الويب، وقد أكدت بعض محركات البحث القليلة تلك الآلية مستندة إلى ما تتبعه قواعد البيانات على الويب. تستخدم تلك الآلية مجموعة من المعاملات مثل < و > و= < و= > التي تلي الكلمات المفتاحية وتسبق تواريخ النشر لوحدات المعلومات. ويعد اعتماد محركات بحث الصور على آلية البحث بحقول الوصف الببليوجرافي يجمع حقول النشر بمكانه وتاريخه، مما يجعل من السهل التخلي عن عزل التغطية الزمنية والمكانية. ويعمل محرك بحث الصور Corbis على استخدام آلية البحث بالحدود المكانية أو الجغرافية بالاعتماد على إحدى القوائم المنسدلة التي تضم مجموعة من الدول هي الولايات المتحدة وألمانيا وفرنسا وأستراليا والمملكة المتحدة واليابان، إضافة إلى اعتماد التغطية الزمنية لملفات الصور وفتراتها التاريخية التي تغطيها في المحتوى وتاريخ التحول إلى الشكل الرقمي.

ويتضح مما سبق أن محركات البحث المحددة قد اعتمدت فقط على آليات بحث أقل مقارنة بمحركات بحث الويب ويمكن إرجاع ذلك إلى أن هذا النوع يعمل في نطاق محدود مع ملفات الصور ومجموعات محدودة العدد من ملفات الصور الرقمية إلى جانب عمومية الموضوعات داخلها، وذلك خلاف محركات بحث الويب التي تتعامل مع أنواع شتى من مصادر الويب مثل صفحات الويب والصور ولقطات الفيديو... وغيرها، مما يتطلب معه الاعتماد على أنواع عدة من آليات البحث بالكلمات المفتاحية وفقًا لطبيعة محرك البحث. فضلا على أن كلا النوعين من محركات البحث لم يعتمد الشكل الكامل من آليات البحث، وهو ما يعد أهم جوانب النقص التي تتصف بها محركات بحث الويب بصفة عامة؛ حيث مازالت تنظر إلى المستوى الأقل مهارة من مستخدمي الويب.

٢/٢/١/٤ البحث بالمسار أو العنوان

يعد البحث بالمسار أو العنوان أول سبل الوصول إلى مصادر المعلومات على الويب، ويعد أسهل السبيلين وصولا إلى المعلومات؛ حيث يعتمد السبيل الثاني على استخدام محركات البحث كوسيط بين المستفيد ومصدر المعلومات الرقمي على الويب. وتخصص الإصدارات المختلفة من متصفحات الويب فراغا يسمى بفراغ البحث بالعنوان والذي يختص بالبحث عن مسار المعلومات التي يطلبها المستفيد

سواء كانت موقعًا قائمًا بذاته أو إحدى صفحات الويب التي تنتمي أو لا تنتمي إلى أحد مواقع الويب. ولا حاجة للمستفيد بأي من المعلومات الأخرى عن مواقع الويب إذا ما علم بمسار تواجده. وتعامل محرك واحد فقط من محركات العينة السابقة وهو Google مع هذه الآلية، في حين أهملته محركات البحث الأخرى، ولم يكن ذلك التعامل في بحث مستقل عن الآليات الأخرى وإنما جاء ذلك كمساعد لأداء مهام الكلمات المفتاحية. وإذا ما وصل الباحث إلى الصور الموضوعية المحددة ما كان له في العودة لهذه الصور سوى البحث بمسارات صفحات الويب المشتملة على تلك الصور المحددة.

٣/٢/١/٤ البحث باستخدام حقول الوصف الببليوجرافي

يقصد بالبحث بالحقول هنا الاعتماد على رموز حقول الوصف الببليوجرافي في استرجاع وحدات من المعلومات تشترك في مجموعة من الحقول مثل تاريخ نشر محدد أو مؤلف واحد أو موضوع مخصص. وتأتي آلية البحث بحقول الوصف تبعًا لاعتماد محركات البحث على وضع تسجيلات ببليوجرافية لما تحويه من مصادر المعلومات سواء أكانت هذه المصادر صورًا أو نصوصًا في أشكال مختلفة. ولا تعتمد محركات البحث على الويب حتى الآن على صياغة شكل واحد للتسجيلة الببليوجرافية في قواعد بياناتها، ويرجع هذا في رأي الباحث إلى أن اعتماد محركات البحث على برنامج المكشف في تكشيف صفحات الويب لا يضع لمثل هذه الصفحات المكشفة سوى الكلمات المفتاحية بما يهمل معه الوصف العام لجوانب صفحة الويب، وما يساعد على ذلك أن برنامج الزاحف يعمل على ضم نسخة من صفحات الويب المضافة إلى قاعدة بيانات محركات البحث، وقد ظهر هذا جليًّا في استرجاع صفحات الويب التي تنتمي إليها الصورة عند عرض التسجيلة الكاملة للصورة.

ولم تستخدم محركات بحث الويب آلية البحث بحقول الوصف الببليوجرافي بالشكل الكامل لاستخدام الحقول الببليوجرافية، وقد جاء الاستخدام الغالب لعناصر البحث الإضافية مثل الحجم واللون والخلفية باعتباره جزءًا مكن من واجهة البحث المتقدم داخل محركات البحث. حيث استخدمت محركات البحث Google, Altavista, Picsearch, Fast آلية الحقول على استحياء من

عدم وجود التسجيلات الكاملة أو الشكل الكامل للوصف الببليوجرافي داخل قواعد بياناتها. وتوارد البحث في الحقول في كل محركات عينة الدراسة على حقول معدودة مثل اللون color or black & white والحجم من خلال القوائم المنسدلة ونوع الملف الذي يحوي الصورة الرقمية سواء كان GIF, JPG, PNG، وقد استخدم Google آلية الحقول عن طرق العبارات التي تشير إلى تحديد النتائج، وما سبق لا يصل إلى ما يهدف إليه الباحث من إيجاد آلية تدعم البحث بالحقول الوصفية للصور الرقمية والتي تساعد استرجاع الصور بالكلمات المفتاحية. يعد النقص السابق لآلية البحث بالحقول أهم نقاط ضعف محركات بحث الصور، ويمكن أن تضع مثل هذه المحركات برنامجًا آخر يعمل على استقطاب صفات وخصائص الصور المادية من ملفات الصور الرقمية بما يسمح بتجميعها فيما بعد واستخدامها متجمعة في شكل تسجيلة ببليوجرافية تكون الصور أحد العناصر المكونة لها. تتيح التسجيلة الببليوجرافية لمحركات بحث الصور النوز بالحسنيين في العمل كتواعد بيانات محكمة تعتمد على العنصر البشري، إضافة إلى السرعة في التحديث والشمول والسعة التي تميز محركات البحث على الويب. وقد أعطى الباحث في الفصل الثالث الشكل المقترح لما يمكن أن تكون عليه التسجيلة الببليوجرافية للصور الفوتوغرافية الرقمية على الويب، بما يمكن أن تستخدم عناصر هذه التسجيلة في استرجاع وحدات المعلومات من الصور الرقمية. ويعتمد Corbis على بعض العناصر البسيطة التي تستخدم كمحدد لبحث مجموعاته مثل عناصر اللون وقطاع البحث مثل etc...news, commercial, entertainment أيضًا الصور داخل أو خارج حق النشر ثم نوع ملف المادة بين صورة فوتوغرافية أو إيضاحات مرسومة.

٤/٢/١/٤ البحث بخصائص اللغة العربية

يعتمد الباحث في دراسة آليات اللغة العربية على محركات بحث الويب فقط، حيث لم تستخدم محركات البحث المحددة آليات التعامل مع اللغة العربية ؛ وذلك يرجع إلى طبيعة بناء المحركات المحددة الذي يتسم بالتحكم البشري والمراجعة المحكمة لمجموعات ملفات الصور المضافة. خلاف ما يكون عليه محرك بحث الويب في إطلاق برنامجه الزاحف crawler في فضاء الويب للتعامل مع شتى أنواع ولغات مصادر الويب.

وتختلف اللغة العربية عن غيرها من اللغات الأخرى في عمليات تنظيم المعلومات، حيث درجة المرونة المطلوبة لتنظيم واسترجاع المعلومات، ومن ثم فإن آليات بحث المعلومات داخل اللغة العربية تختلف عن تلك المستخدمة في اللغات الأخرى مثل اللغات اللاتينية على أنواعها المختلفة. ودخل استرجاع اللغة العربية إلى عهد جديد معتمدا على تقنيات التحليل الصرفي والمعجمي لألفاظ وتراكيب اللغة العربية، وانتفت مع هذه التقنيات معظم مشكلات الاسترجاع الآلي للمعلومات في اللغة العربية. اعتمدت بعض محركات البحث العامة باللغة العربية على مثل تلك التقنيات، ومن أهمها محرك Ajeeb الذي انعكس استخدام تقنيات التحليل الصرفي به في آليات البحث بالمعنى والمشتقات والجذر والمترادفات.

وحري بالذكر هنا أنه لا توجد محركات بحث للصور تعمل باللغة العربية من عينة الدراسة سوى Google, Picsearch, Altavista, Fast. وسيتم بالاعتماد على تلك المحركات دراسة خصائص استرجاع الصور في بيئة اللغة العربية. ويوضح الجدول رقم (٤-٣) آليات البحث في اللغة العربية وتطبيقها في محركات بحث الصور على الويب:-

الجدول رقم (٤-٣) يوضح آليات بحث اللغة العربية في محركات بحث الصور

	محركات بحث الويب						محركات بحث محددة	
	Alta- Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
البحث التطابق	✓	--	✓	✓	x	--	--	--
البحث غير المطابق باللواصق	x	--	x	x	x	--	--	--
البحث بجذر الكلمة	x	--	x	x	x	--	--	--
البحث بالتشكيل	x	--	x	x	x	--	--	--
البحث الحر	x	--	x	x	x	--	--	--
البحث بالترادف	x	--	x	x	x	--	--	--

تابع ـ الجدول رقم (٤-٣) يوضح آليات بحث اللغة العربية في محركات بحث الصور

	محركات بحث الويب						محركات بحث محددة	
	Alta- Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
البحث بالمتضادات	x	--	x	x	x	--	--	--
البحث بالمعاني	x	--	x	x	x	--	--	--
البحث بالترجمة	x	--	x	x	x	--	--	--
التعرف إلى الأخطاء العربية الشائعة	x	--	x	x	x	--	--	--

وتحتاج اللغة العربية إلى مجموعة من آليات البحث كتلك المشار إليها في معالجة اللغة الطبيعية في النصوص العربية، ولم تزل محركات بحث النصوص العربية تعمل بآليات بحث عامة تتوافق مع الكلمات المفتاحية في مختلف اللغات الأخرى؛ مثل آليات المنطق البوليني، والبحث الحر، والتطابق، والبحث بالتقارب... وغيرها من الآليات الأخرى. ولقد عملت المحركات العربية حديثًا على استخدام التقنيات المساعدة في بحث اللغة العربية مثل تقنيات التحليل المعجمي والتحليل الصرفي، ولا يمكن بحث النصوص العربية المستقلة أو التابعة لوسائط الصور والصوت دون الاعتماد على التقنيات السابقة. لقد خلت محركات بحث العينة السابقة حتى المستخدم للعربية منها من آليات بحث تعمل خصيصًا لبحث وحل مشكلات استرجاع المعلومات من النصوص العربية، نظرًا لخلوها من تقنيات التحليل الصرفي والتحليل المعجمي. واعتمادًا على التقنيات السابقة يمكن إضافة مجموعة من آليات البحث إلى محركات بحث الصور لدعم اللغة العربية، وتتمثل هذه الآليات فيما يلي :-

١. البحث بالتطابق
٢. البحث غير المطابق باللواصق
٣. البحث بجذر الكلمة

٤. البحث بالتشكيل
٥. البحث الحر
٦. البحث بالترادف
٧. البحث بالمتضادات
٨. البحث بالمعاني
٩. البحث بالترجمة
١٠. التعرف إلى الأخطاء العربية الشائعة
١١. طرح سؤال للبحث : قائمة توقف عن الأحرف والأدوات

وقد أوضح الملحق رقم ٦ طبيعة النتائج التي يتم استرجاعها بآليات البحث العربية في محركات البحث.

٢/٤ خصائص محركات بحث الصور

يتناول هذا القسم تحليل خصائص محركات بحث الصور من حيث الوصف الببليوجرافي، ومستويات التكشيف، ولغات الوصف كما تم من اختبار محركات البحث الخاضعة لعينة الدراسة.

١/٢/٤ الوصف الببليوجرافي للصورة

تتساوى محركات البحث العامة والمتعددة على الويب مع محركات البحث المحددة نوعيًا أو موضوعيًا في الاحتياج إلى شكل مقنن من أشكال ومعايير الوصف الببليوجرافي. وتحتاج محركات بحث الصور بطبيعة الوسيط إلى العناصر المادية أو الشكلية لاستخدامها في استرجاع الصور المتكاملة مع العناصر الموضوعية المتمثلة في الكلمات الكشفية. ولا تحتوي محركات بحث العينة السابقة جميعها على العناصر المقننة لوصف وسيط للصور الفوتوغرافية العامة الرقمية، وقد عكست آليات البحث المدروسة في محركات بحث الصور انتفاء البحث بحقول الوصف الببليوجرافي في جميع محركات بحث العينة. وتتميز قواعد بيانات الصور (محركات البحث المحددة)عن نظيرتها في محركات بحث الصور على الويب بتقديم تسجيلات ببليوجرافية في محاولة للتغلب على مشكلات استرجاع الصور،

وهو ما يحاول هذا الفصل إضافته إلى محركات بحث الصور في التعامل مع الملفات الرقمية للنصوص أو الصور واستخراج حقول الوصف.

وبالاعتماد على الشكل المقترح للتسجيلة الببليوجرافية للصور الرقمية على الويب في الفصل السابق المتمثل في المعيار المقترح لوصف الصور الرقمية لـدبلن كور، يمكن قياس وفحص محركات بحث العينة ومن ثم التعرف إلى الأشكال البسيطة من الوصف الببليوجرافي لمجموعاتها من الصور الرقمية المتمثل في التسجيلة النهائية للصورة؛ وهو ما يعني به الباحث الشكل الكامل من الوصف الببليوجرافي لوسيط المعلومات الذي يعرضه نظام الاسترجاع إذا ما وقع اختيار المستفيد على هذا الوسيط. ويمثل الجدول رقم (٤-٤) جانب الوصف الببليوجرافي لمحركات بحث الويب :-

الجدول رقم (٤-٤)

يوضح حقول الوصف الببليوجرافي التابعة لتسجيلات الصور الرقمية

	محركات بحث الويب						محركات بحث محددة	
	Alta-Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
العنوان :	✓	×	✓	×	✓	×	✓	✓
الفنان، المصور (المؤلف) :	×	×	×	×	×	×	✓	✓
الموضوع :	×	×	✓	×	×	×	×	✓
وصف الموضوع (المحتوى)	×	×	✓	×	×	×	×	×
تاريخ التقاط الصورة :	×	×	×	×	×	×	✓	✓
تاريخ صنع الصورة الرقمية	×	×	✓	×	×	×	✓	×
تاريخ وضع التسجيلة :	×	×	✓	×	×	×	✓	×
الناشر :	×	×	×	×	×	×	✓	✓
شكل ملف الصورة :	✓	×	✓	×	✓	×	×	×

تابع - الجدول رقم (٤-٤)

يوضح حقول الوصف الببليوجرافي التابعة لتسجيلات الصور الرقمية

	محركات بحث الويب						محركات بحث محددة	
	Alta-Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
حجم ملف الصورة :	✓	x	✓	x	✓	✓	x	x
أبعاد الصورة :	✓	x	✓	x	✓	x	✓	x
اللون :	✓	x	x	x	✓	x	✓	x
مسار الملف :	x	✓	x	✓	✓	✓	x	x
مصدر الصورة على الويب :	✓	✓	x	✓	✓	✓	x	x
لغة ملف الصورة	x	x	x	x	x	x	x	x
تغطية الصورة · المكان :	x	x	x	x	x	x	✓	✓
حقوق نشر الصورة :	x	x	x	x	x	x	✓	x

يوضح الجدول رقم (٤-٤) الكثير من الجوانب العامة التي يمكن تبيانها فيما يلي:-

- يعد محركا البحث Visoo, Ditto من أفقر محركات بحث الصور لحقول الوصف الببليوجرافي، وهذا يرجع كما لاحظ الباحث إلى عدم احتواء محركي البحث لتسجيلات ببليوجرافية عن مجموعاتها من الصور الرقمية؛ إنما ترتبط هذه المحركات بصفحة الويب التي تحوي الصورة المحددة، مما يعمل على الانتقال إلى صفحة ويب أخرى بعيدة عن نطاق محرك بحث الصور، ويعمل محرك Visoo على ربط اللقطات المعروضة من الصور الرقمية بحجم أكبر لعرض الصورة بدون اتباعها بأية بيانات وصف ببليوجرافية، على أنه يخصص أحد بيانات عرض النتائج المتمثل في مسار الصورة للربط مع صفحة الويب مصدر الصورة. والحقول التي دلت عليها العلامات في الجدول لهذين المحركين إنما تأتي في واجهة العرض وليس في تسجيلة ببليوجرافية منفصلة.

يقدم محركا البحث Google، Picsearch إطارًا آخر مختلفًا عن إطار عرض التسجيلة الببليوجرافية وهو الإطار الذي يضم صفحة الويب التي تحوي الصورة الرقمية الأصلية، إلا أن محرك البحث Google يعتمد كلية في التعريف بالصورة الرقمية على إطار صفحة الويب بدلا من استخدام الحقول الببليوجرافية، فكان من الطبيعي أن يحتوي محرك بحث الصور Google على عنصرين أو حقلين من حقول الوصف فقط هما مسار ملف الصورة الرقمية، ومسار صفحة الويب مصدر الصورة.

تستخدم بعض محركات بحث الصور البحث بحقول أو عناصر وصف ببليوجرافي، على الرغم من أنها لا تستخدم تلك العناصر في صياغة التسجيلة الببليوجرافية. ويمثل ذلك محرك البحث Google حيث يعرض مجموعة من الحقول المتاحة للبحث مثل نوع ملف الصورة، واختيار حقل اللون، وحجم الصورة.

يتصف محرك بحث الصور Corbis بالغالبية من صفات محركات البحث الأخرى؛ إلا أن كفاية عناصر البحث وحقول التسجيلة الببليوجرافية قد عكست الدور الذي يؤديه العنصر البشري في تغيير كفاية البحث في محركات الصور الرقمية، من خلال تقديم خصائص لوصف المحتوى الموضوعي للصور تتوقف على البعد الثالث للصورة، وهو ما يعجز عنه حتى الآن برامج الزاحف في محركات بحث الويب.

يعد محرك البحث Freefoto من أفقر محركات عينة الدراسة كاملة في الاعتماد على اختزان ملفات الصور الرقمية في بيئة الوصف الببليوجرافي على الرغم من كونه الأقرب إلى التعامل مع الضبط والتقنين لمجموعاته المحدودة. ويأتي محرك البحث Corbis ليكون أفضل محركات العينة في الاعتماد على عناصر وصف الصور الرقمية. وعلى الرغم من عدم إظهار هذا الأمر في الشكل النهائي لصورCorbis إلا أن خيارات البحث المتقدمة لم تكن تركز على آليات البحث التقليدية مثل المنطق البوليني أو البحث الحر، وإنما اعتمدت على البحث بحقول مثل التغطية والبنية والمصور والمحتوى..، مما عكس للباحث الحقول التي يتم إدخالها عن كل صورة رقمية مكونة التسجيلة النهائية، ثم إخضاع كل هذه الحقول لعمليات البحث أمام المستخدم.

ونستنتج مما سبق أنه لم يستخدم أي من محركات البحث السابقة مفهوم التسجيلة الببليوجرافية في المعنى الكامل لها سوى المحاولة التجارية لدى Corbis، غير أن التعامل مع حقول البحث قد حدث اجتهادا من محركات بحث الويب والمحددة لاستكمال عناصر البحث المتقدم بعيدًا عن كونها عناصر للبحث في حقول تسجيلة تم بناؤها مسبقًا.

وتعكس دلالات النتائج السابقة تأخر محركات البحث في التعرض إلى نهج معياري أو مقنن تستخدمه محركات بحث الصور في التعامل مع أوعية المعلومات النصية أو أوعية المعلومات ذات الوسائط الأخرى، ولا يتوقف هذا الأمر عند وسيط معلومات الصورة فقط وإنما يتعدى إلى أوعية الفيديو والصوت والرسوم المتحركة. وتأتي أهمية التسجيلات الببليوجرافية من احتوائها على مجموعة كبيرة من مفاتيح الوصول إلى تسجيلات ووحدات الصور الرقمية الطبيعية.

وجاءت محركات البحث Corbis, Fast, Altavista, Picsearch على الترتيب كأفضل محركات عينة الدراسة في التعامل مع مفهوم ومحتوى التسجيلة الببليوجرافية. حيث اعتمدت هذه المحركات على استخدام حقول ثابتة، ثم تفريغ بيانات ملفات الصور الرقمية داخلها، أو إظهارها فقط أمام مستخدم الويب في مرحلة البحث كما فعل Corbis، وقد وصل الأمر إلى استخدام محرك Fast لشكل مدخل التسجيلة في البدء بعنوان الصورة ثم وصف العناصر الموضوعية في جملتين أو أكثر، وبالمثل كان محرك Altavista ثم اعتماد محرك Picsearch على استخدام عناصر الوصف على الرغم من عرض الصورة الرقمية مع صفحة الويب في إطار واحد. وكل ذلك يعكس أنه يمكن لمحركات بحث الويب ضبط وتقنين اختزان ملفات الصور داخلها، ذلك بالاعتماد على إدخال سياسات وصف وتكشيف مقننة لاقتباس وتفريغ بيانات الصور الرقمية من صفحات الويب. ومن ثم تصبح محركات بحث الويب، على غير المتوقع، أكثر كفاية من محركات البحث المحددة في التعامل مع عناصر الوصف الببليوجرافي.

٢/٢/٤ مستويات التكشيف

لقد آثر الباحث في هذه الخاصية الجمع بين نوعي محركات البحث، وذلك للاعتماد كلية على تحليل المصادر النظرية التي تصف عملية التكشيف داخل

محركات البحث دون الاعتماد على المشاهدة أو الملاحظة كباقي الخصائص السابقة.

وتختلف وسائط معلومات الصور الرقمية عن الوسائط النصية التقليدية والرقمية من حيث عمليات التحليل الموضوعي. ويؤثر وسط حمل المعلومات في الصور الرقمية والمبني على الرؤية في عمليات التحليل الموضوعي المتمثلة في عمليات التكشيف للصور على مستوى التحليل الموضوعي للصور الرقمية. وإن ما سبق من اختلاف وسط حمل المعلومات في الصورة عنه في النص يعكس مستويات مختلفة من المعلومات يمكن للصورة أن تحملها، وتتجلى هذه المستويات أمام المستفيدين بالنظر إلى وسيط الصورة، وتؤثر أيضًا اختلافات الرؤى إلى اختلاف المحتوى الموضوعي وتختلف على أثره الكلمات المفتاحية والواصفات. وتتمثل المستويات المختلفة للمعلومات داخل الصور إلى ثلاثة مستويات؛ الأول ويتعلق بمجموعة العناصر الأساسية التي تراها أعين المستفيدين أول ما تقع على الصورة، أما الثاني فهو مجموعة العناصر التي تمثل الخلفية لمجموعة العناصر السابقة، وأما المستوى الثالث فهو مجموعة المعاني والمفاهيم أو الأفكار أو المعاني التي لا تمثلها عناصر مادية تراها العين في الصورة، وهو ما يطلق عليه البعد الثالث في رؤية أوعية المعلومات المرئية.

ويمكن التعامل مع المستويات الثلاثة السابقة في تكشيف الصور الرقمية إذا ما أخذت محركات البحث على عاتقها استخدام العنصر الإنساني والبشري في التعرف إلى المعاني والمفاهيم داخل الصورة. ويمكن أن يتأتى ذلك من خلال أحد العاملين وهما؛ أولا الاعتماد على التكشيف الآلي للصور الرقمية داخل قاعدة بيانات محرك البحث مع إمكانية إدخال التعديلات النهائية بشريًا، ثانيا يمكن لمحركات البحث أن تطرح مجموعاتها المضافة حديثا بواسطة برنامج الإضافة الزاحف أو العنكبوت أمام مستخدمي الصور على الويب لاقتراح المصطلحات التي تلائم هذه الصور في صورة تفاعل بين محرك الصور ومستخدمي مجموعات الصور، وبالاعتماد على المصطلحات الأكثر تواترًا وتكرارًا لكل صورة من الصور الرقمية يتم الوصول إلى الكلمات المفتاحية المعبرة عن الأبعاد الثلاثة لتكشيف الصورة.

إن محركات البحث جميعها تعمل الآن على إجراء عمليات التكشيف معتمدة على النص المصاحب للصور الرقمية خاصة لتواجدها دائما كجزء مكون لصفحة

الويب التامة، وقد اتخذت محركات بحث الصور التكشيف الكامل لكل كلمات متن صفحات الويب مضمارًا للتسابق، وأشار Sullivan, Danny إلى أن هذا بجانب عمليات لتكشيف الصور التي تعتمد على حقول ورموز اللغة المعيارية لصفحات الويب HTML meta tags إلى جانب الاعتماد على حقل الكلمات المفتاحية HTML Kew Word.(١)

وتعتمد بعض محركات بحث الصور الأخرى على الكلمات ذات المواقع المهمة لعرض الصفحة المصدرية مثل كلمات العنوان Title، وكلمات مسار الصفحة URL، إلى جانب الكلمات المحيطة بإطار الصورة والواصفة لمحتواها. وقد دفع الاعتماد على كلمات النص والحقول المختلفة لملفات الصور باللغة المعيارية في تكشيف الصور الرقمية دفع بعض محللي صفحات الويب إلى تطوير شكل النص وحقول اللغة المعيارية للعمل بكفاية أكثر على تقديم الكلمات المفتاحية المعبرة عن المحتوى؛ وقد طرحت Larisa Thomason عناصر تطوير ملفات الصور وصفحاتها في(٢):-

١. الكلمات الواصفة في العنوان Descriptive image file names، حيث يشمل العنوان كل الكلمات المعبرة عن عناصر الصور أو الموضوع الرئيس للصورة، بدلا من الكلمات غير المعبرة عن الموضوع.

٢. الاهتمام بحقول الكلمات المفتاحية وحقول الوصف ALT text descriptions، حيث يضع مصمم الصفحة المصدرية وملف الصورة الرقمية كل الكلمات التي تصف محتوى الصورة وبأبعاده الثلاثة، والعمل على تركيز الكلمات المعبرة فقط.

٣. الاهتمام بالكلمات المحيطة بالصورة الرقمية surrounding content، حيث توضع الصور في مجموعة من الكلمات أو السطور الواصفة لمحتوى الصورة ويقترح الباحث أن تخصص مثل هذه السطور والجمل إلى التحليل الموضوعي باستخدام النص الشارح والذي سبق الحديث عنه في أشكال لغات تكشيف الصور الرقمية.

---

(١) Sullivan, Danny. Death Of A Meta Tag, Jupiter media Corporation, 2002, visited 25/ 12/1/2002, http://searchenginewatch.com/subscribers/articles/02/10-meta.html

(٢) Thomason, Larisa. Promotion Tip: Search Engines Get The Picture, 2002, visited 25/ 12/2002, http://www.netmechanic.com/news/vol5/promo_no3.htm

وتتبنى محركات البحث اتجاه التعامل مع الصورة من خلال النص المصاحب للصورة، وهو ما يجعل التأكيد على تحليل الصورة ومستويات تكشيفها أمرًا ضروريًا. وعلى الجانب الآخر، فإن ثمة تقنية أخرى لتحليل الصورة تحاول الدخول إلى عمل محركات بحث الصور وهي تحليل المحتوى بعناصر اللون والشكل والبنية فيما يطلق عليه الاسترجاع المبني على المحتوى.

ويهدف هذا الجزء من الدراسة التجريبية إلى التعرف إلى طرق التكشيف لدى محركات بحث العينة السابقة لمحركات بحث الصور، ويوضح الجدول رقم (٤-٥) مصادر محركات بحث الصور التي تمدها بالكلمات الواصفة ومداخل الاسترجاع في مرحلة البحث، ويعتمد الباحث في تحليل الجدول رقم (٤-٥) على الدراسات والمقالات التي تناولت التكشيف داخل محركات بحث العينة، وهذا يرجع إلى إخفاء محركات البحث لسياساتها في العمليات الفنية التي تقوم بها. بالتطلع إلى خصائص الكلمات المفتاحية المصاحبة للصورة إذا ما تم استرجاع تلك الصور، ويتم في هذا الشأن الاعتماد على ثلاثة مجموعات من الكلمات المفتاحية تمثل المستويات الثلاثة لتكشيف الصور.

الجدول رقم (٤-٥) يوضح مصادر تكشيف الصور داخل محركات البحث

		محركات بحث الويب						محركات بحث محددة	
		Alta-Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
حقول ملف الصورة والصفحة		--	--	--	--	--	--	✓	--
النص الكامل		✓	✓	✓	✓	✓	✓	--	✓
مواقع الكلمات داخل الصفحة	العنوان	--	--	--	✓	--	--	--	--
	حول الصورة	✓	--	--	✓	✓	--	--	✓
	المسار	--	--	--	--	--	--	--	--

ويتضح من الجدول رقم (٤-٥) مصادر الحصول على الكلمات الكشفية للتعبير عن محتوى الصور داخل قاعدة بيانات محرك البحث، وقد ظهر جليًا من

الجدول الاطراد في تعامل محركات البحث مع الصور مثل النصوص في معالجتها، حيث تعاملت المحركات السابقة مع تكشيف المتن الكامل للصفحة المصدرية Host Page لإطار الصورة. وقد شذ عن ذلك محرك البحث Corbis، حيث يتعامل هذا المحرك مع تسجيلات غير كاملة للصور الرقمية تتكون من بعض حقول الوصف المادي والموضوعي، وذلك لما له من طبيعة تكوين قواعد البيانات وليس محركات البحث، حيث تعالج الصورة من قبل مجموعة من المكشفين القائمين على اختيار الكلمات المفتاحية الملائمة للتعبير عن المحتوى. وتستخدم محركات البحث في حالة الاعتماد على التكشيف الكامل لنص الصفحة كل عناصر بناء الصفحة مثل تكشيف العنوان وتكشيف حقول الكلمات المفتاحية والمسار. وقد اعتمد الباحث في الوصول إلى مصادر تكشيف محركات البحث لصفحات الويب إلى المقارنة التي عقدها RICHARD EINER PETERSON عن مجموعة من السمات المشتركة بين محركات المقارنة.(١)

٣/٢/٤ لغات الوصف المستخدمة في التكشيف

تعتمد محركات بحث الصور في تكشيف مجموعاتها على المتن الأصلي لصفحة الويب أو تلك الرموز التي يضعها مصممو صفحات الويب داخل ما يعرف بالـ HTML meta tags. وعلى ذلك فإن مستويات الوصف للمحتوى الموضوعي للصورة تعتمد كلية على وصف الموضوع داخل صفحة الويب المكون من مجموعة متكاملة من الجمل والعبارات داخل صفحة الويب. ولا تتعامل محركات بحث الصور مع الصور الرقمية التي تصاحب الصفحات النصية، وإنما تتعامل مع كل صور الويب التي يمكن أن تمتد إليها قبضة برامج الزاحف والعنكبوت. ولا تخلو الويب من الصور المنفردة التي لا تصاحب جملًا أو عبارات واصفة للموضوع. ويقترح الباحث التعامل مع الصور الرقمية من خلال ثلاث لغات وصف للتعبير عن محتوى الصورة المرئي وتحويله إلى وسط النص، وتتركز هذه اللغات أولا في المصطلحات المنفردة وتبنى على عرض محتويات الصورة هنا باختلاف مستويات التكشيف (العناصر الأساسية والخلفيات والبعد المعنوي) في واصفات أو مصطلحات منفردة. وتأتي

---

(١) PETERSON, RICHARD EINER. Eight Internet search engines compared, first Monday, 1997, visited at 21/6/2003, http://www.firstmonday.dk/issues/issue2_2/peterson/

العبارات لغة تعبير ثانية تبنى على استخدام عبارات غير مرتبة أو مترابطة تعبر كل منها على أحد العناصر الموضوعية في الصورة الرقمية، على أن تشمل هذه العبارات عناصر المستويات الثلاثة لتكشيف الصورة الرقمية. أما لغة التعبير الثالثة فهي الجمل أو المترابطة أو النص الشارح : وهو ما تعتمد عليه محركات البحث الحالية في تكشيف الصور، وتتمثل الجمل المترابطة في النص المصاحب للصورة. وقد تعرض الفصل الثالث بالتفصيل لذلك في مناقشة مستويات تكشيف الصور الرقمية في إطار المعالجة الفنية للصور الرقمية.

وتتضح مختلف لغات الوصف السابقة عند قيام محركات بحث الصور في أداء مهامها لاسترجاع الصور المحددة، ويتحدد الأسلوب المتبع لمحرك البحث في النظر إلى التسجيلة النهائية، وما يصاحبها من وصف للموضوع سواء باستخدام المصطلحات المنفردة أو استخدام العبارات أو جمل الوصف (النص الشارح). ويوضح الجدول رقم (٤-٦) الأساليب المتبعة في محركات بحث العينة للتعرف إلى محتويات الصور وتنظيمها، والتي تستخدمها على الجانب الآخر في إتاحتها للبحث واستدعاء الصور. حيث يمثل المصطلحات المنفردة كل من مسار الملف أو مسار صفحة الويب (مصدر الصورة)، وعنوان الصورة، والكلمات المفتاحية، ورموز Meta Tags. ويمثل العبارات الجمل غير التامة لوصف الموضوع للتعرف إلى المحتوى. أما الجمل الشارحة (النص) فيمثلها في تنظيم واسترجاع الصور داخل قواعد البيانات تلك النسخ من الصفحات المصدرية للصور الرقمية. ويكون الجدول كما يلي:-

ويهدف الباحث إلى إبراز لغات الوصف وليس لغات التكشيف المنطوية على اللغة المقيدة واللغة الطبيعية. وتعتبر لغات الوصف هذه الأقسام الأساسية للغة التكشيف الطبيعية التي تتناسب أكثر مع الصور الفنية أو العامة أو الطبيعية، بعكس ما يكون في الصور العلمية التي تتناسب أكثر مع اللغات المقيدة؛ نظرًا للتخصص الموضوعي الذي تنتمي إليه مجموعات الصور. ويناسب استخدام النص الشارح التعامل مع الصور وتحويل المحتوى من الوسط المرئي إلى وسط النص؛ نظرًا لما يعتري الصور من تعدد عارم في الموضوعات سواء التي تدل عليها عناصر الصورة ذاتها، أو تلك التي تدل عليها الصورة بصفة عامة؛ نظرًا لطبيعة الفترة الزمنية التي تنتمي إليها أو طبيعة البيئة الجغرافية والمكان الذي

الجدول رقم (٤-٦)

يوضح لغات التعبير عن المحتوى الموضوعي للصور لمحركات البحث

		محركات بحث الويب						**محركات بحث محددة**	
		Alta-Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
المصطلحات المنفردة	**المسار**	✓	✓	--	✓	✓	✓	--	--
	الكلمات المفردة	--	--	--	--	--	--	--	--
	عنوان الصورة	✓	--	✓	--	✓	--	✓	✓
	الرموز المعيارية	--	--	--	--	--	--	--	--
العبارات		--	--	✓	--	--	--	--	✓
الجمل أو النص الشارح		--	--	--	✓	✓	--	--	--

تنتمي إليه، حيث إن المصطلحات المنفردة والعبارات في اعتمادها على مبدأ الإحاطة والشمول تحتاج إلى أعداد كبيرة من وحدات الوصف وقد لا يتمكن المكشف من التعبير عن ذلك. إن استخدام لغات الوصف في التعبير يهتم أكثر ما يهتم بقواعد بيانات الصور ذات البناء البشري اليدوي ثم الاسترجاع الآلي؛ حيث تعتمد محركات البحث على متن الصفحة أو حقول بناء ملف لغة الترميز المعيارية. وإذا ما كان الأمر كذلك في محركات البحث وقواعد البيانات، فإن الباحث يقترح إدخال حقل آخر داخل ملفات الويب يتسع للجمل أو النص الشارح لمحتوى الصور الرقمية، ويمكن على الجانب الآخر استخدام حقل Key Words في بداية الملف لوصف محتوى الصور. ويأتي كل ما سبق من طبيعة محركات البحث في التكشيف بما هو قائم وليس بإضافة كلمات أو جمل يختص بها محرك البحث لمجموعاته من الصور الرقمية.

ويتضح من الجدول رقم (٤-٦) الحالات المختلفة التي توضع لوصف الموضوع داخل محركات بحث الصور. ويعتبر محرك البحث Picsearch أفضل المحركات في اعتماده على عناصر مختلفة بين عنوان ومسار الصورة والمتن الأصلي للصفحة المصدرية، وهو ما يعطي إمكانية التعرف إلى مختلف الموضوعات التي تشملها الصورة سواء في مرحلة التكشيف أو في مرحلة البحث والاستدعاء. على الجانب الآخر فإن محرك البحث Altavista أقل محركات البحث تعددًا لمصادر التكشيف والتعرف إلى محتويات الصورة في عرض الصور؛ إلا أن واقع عمل محركات بحث الصور جميعًا يعتمد على نصوص الصفحات المصدرية، ولا يعتبر ذلك سمة خاصة لمحرك Altavista وإنما سمة مختلف محركات البحث التي تقتصر في عرض تسجيلاتها على أقل العناصر بتركيزها على نص الصفحات المصدرية؛ وهو ما دلت عليه نتائج البحث عن صور تتعلق بمصطلح Rose ؛ حيث تعدت النتائج المستدعاة عشرات الآلاف من الصور الرقمية من مجموعات محركات البحث مما يدل على الاستدعاء العالي المبني على الشمول والإحاطة في التكشيف.

٤/٢/٤ واجهات محركات بحث الصور

تمثل واجهات محركات البحث بصفة عامة، ومحركات بحث الصور بصفة خاصة، السبيل الأول للمستخدم للتعرف إلى نظام الاسترجاع داخل محركات البحث، وتتمثل أهمية الواجهات باعتبارها أداة تعريف لمحتويات محرك البحث، كما أنها مؤشر على البساطة أو التعقيد في التعامل مع عملية الاسترجاع في محركات البحث. وتزداد هذه الأهمية في محركات بحث الصور وهي التي تتعامل مع الوسائط المرئية (الصور). وما يغلب على واجهات محركات بحث الصور أنها تخدم البحث والاسترجاع في قاعدة بيانات صور المحرك؛ باحتوائها على مجموعة من أفضل الصور داخل قاعدة البيانات. كما تحتوي أيضًا على مجموعة من المصطلحات التي تكون في غالب الأمر الأقسام العامة لموضوعات الصور العامة،أو مجموعة من الأقسام العامة التي تندرج أسفلها مجموعات أكثر تخصصًا من المصطلحات لخدمة الاسترجاع بتلك المصطلحات وهي ما يتضمنه أسلوب التصفح في محركات بحث الصور. ويشمل الحديث التالي عن واجهات محركات البحث في تحليل تلك الصفحات والواجهات التي تخدم البحث المتقدم للصور، ونقاط القوة والضعف التي يخلفها تصميم هذه الصفحات وأثرها على استرجاع الصور.

وتختلف واجهات البحث بين محركات الصور وبعضها البعض، حيث يعتمد معظمها على وضع الصور ذات الموضوعات المهمة أو الجذابة للمستفيدين بشكل عام. وتبعًا لعينة البحث السابقة، فإن هذه العينة اختلفت في التعامل مع هذا الجانب؛ فمنها من ضم الواجهات الجذابة كثيرة الألوان ومنها ما اعتمد على الواجهات البسيطة المقتصرة على فراغ البحث وبعض المحددات الصعبة في تحديد النتائج المطلوبة. وتغلب طبيعة محرك البحث على واجهته في كل محركات البحث؛ فإن نظم استرجاع الصور التي تهدف إلى الربح تعمل على وضع شروط حق النشر إضافة إلى تحديد خدمات المحرك التي يمكن أن يقدمها إلى مستخدميه. ويوضح الجدول رقم (٤-٧) الجوانب التي يمكن أن توجد بها واجهات محرك البحث:-

الجدول رقم (٤-٧) يوضح عينة البحث مع عناصر واجهات البحث

		محركات بحث الويب						محركات بحث محددة	
		Alta-Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
عرض بعض الصور		--	✓	--	--	✓	--	✓	--
واجهات بسيطة		✓	✓	--	✓	✓	✓	--	✓
تعدد واجهات البحث		✓	--	✓	✓	✓	--	✓	--
شكل آليات البحث	قوائم	--	--	--	✓	--	✓	✓	--
	ترميز	✓	✓	✓	--	✓	--	--	--
المساعدة لواجهة البحث		✓	--	--	--	✓	--	--	--
عناصر التفاعلية		--	--	--	--	--	--	--	--

ويشير الجدول رقم (٤-٨) إلى مجموعة العناصر التي تكون واجهات محركات بحث الصور، وتختلف هذه العناصر فيما بينها كعناصر مؤثرة في استرجاع الصور وتبسيط البحث. فقد تضمن العنصر الأول مجموعات الصور التي يمكن أن تضمها واجهة البحث، وأن واقع بحث الصور داخل نظم استرجاعها يعكس

أن هذه الصور ليست ذات تأثير في استرجاع الصور، وباختبار عناصر الانتظام داخل مجموعة الصور، فإنها تهدف فقط إلى جذب انتباه المستفيد أو لإضفاء شكل مميز لمحرك بحث الصور. إن مجموعات الصور التي يمكن أن تضمها واجهة البحث هي التي تخدم استرجاع الصور العلمية في حالة تقسيم الموضوعات العلمية إلى تفريعاتها المتخصصة. وقد حملت ثلاثة واجهات من عينة البحث صورا بداخلها، وهي Picsearch, Corbis, Ditto. وعلى ذلك فإن تعدد الصور في واجهة البحث هو إحدى نقاط القوة في عرض التقسيم الموضوعي لمحرك البحث باستخدام الصور، أو تكون خلاف ذلك بعرض الصور لجذب الانتباه مما يكون له أثر سلبي في عرض واجهة محرك البحث.

وتسبب مشكلة الانتقال من صفحة إلى أخرى داخل محركات البحث مشكلة قد لا يستطيع المستفيد الوصول إليها، فضلا عن الوقت المضاف إلى عملية البحث. وعلى هذا فإن شطر البحث داخل النظام إلى واجهة بحث بسيطة مع أخرى تحمل البحث المتقدم يمكن أن لا يفطن إليه الكثير من مستخدمي الويب، كما يكون له تأثير سلبي على الاستفادة بكل ما يقدمه محرك البحث من آليات بحث الصور. وقد اعتمدت أربعة محركات من العينة وهي Picsearch, Google, Corbis, Fast على أكثر من واجهة لبحث الصور تحمل الأولى آليات عامة ثم تحمل الأخرى الآليات الأكثر تقدمًا. إن احتواء واجهة بحث واحدة لمختلف آليات البحث داخل قاعدة البيانات، يمكن المستفيد من التعرف إلى مختلف الآليات المستخدمة مما يترتب عليه التعامل مع مختلف هذه الآليات بحسب الحاجة الموضوعية لمستخدمي محرك بحث الصور.

إن أهم ما يميز صفحات الويب هو كثرة الألوان والأشكال المرئية التي تملأ صفحات الويب؛ إلا أن ذلك قد يبعد بالمستفيد عن التركيز الكامل في التعرف إلى وحدات وآليات البحث. وعلى ذلك فإن عرض واجهات البحث بشكل مبسط يجعل من السهل التعرف إلى محتويات الصفحة المختلفة من توجيهات وآليات البحث، إن البساطة التي يعنيها الباحث هي التنظيم في عرض الرسائل الموجهة للمستفيد والتي بطبيعة الأمر تساهم في تيسير عملية البحث، وهي ما يمكن أن توجد حتى في ظل الرسائل النصية الكثيرة في واجهة البحث. ويدخل في إطار التيسير الشكلي المتبع في بحث الصور ومحركات البحث بصفة خاصة، هو شكل آليات البحث؛ فإن

بعضًا من محركات البحث تعتمد على القوائم المنسدلة والبعض الآخر يعتمد على تعيين آليات البحث المطلوبة بوضع الرموز المحددة لها. ويعد الشكل الثاني أقرب إلى تسهيل الرؤية واختيار آليات البحث، هذا فضلا عن ترك الكثير من المستفيدين لمعظم القوائم المنسدلة دون اكتشاف ما تحويه هذه القوائم. وقد اتبعت أربعة من محركات البحث القوائم في حين اتبع اثنان التحديد بالرموز مثل Fast, Picsearch في حين اكتفت محركات مثل Ditto بواجهة البحث التي تحمل فراغ الكلمات المفتاحية فقط دون آليات أخرى، وهو ما يعكس غياب الدقة في بحث قاعدة بيانات محرك الصور.

ويأتي الأمر الأخير في واجهات بحث الصور والمتعلق بتلميحات المساعدة التي توفر للمستفيد المعرفة التامة بقواعد البحث وأنواع آليات البحث المستخدمة داخل نظام استرجاع بعينه، وليس بالخفي أن السواد الأعظم من مستخدمي الويب لا يعرفون الأنواع المختلفة من آليات البحث داخل محركات البحث، هذا فضلا عن استخدام أشكال مختلفة بين محركات البحث للتعبير عن آليات بحث واحدة مثل استخدام (-، +) للتعبير عن معاملات المنطق البوليني (and, not)، مما يكن له أكبر الأثر في ضياع كم من التحقيق أمام المستفيد.

ويخلص الباحث من الحديث عن عناصر واجهات البحث إلى أن واجهات بحث الصور يجب أن تتميز بعناصر شكلية تساعد الباحث في إجراء عملية البحث مستخدمًا مختلف الإمكانات والآليات التي يوفرها نظام استرجاع الصور. ولا بد أن تتميز واجهة البحث بخلوها من الأعداد الكبيرة من الصور التي لا تهدف إلى عرض تقسيم موضوعي؛ إلا أن الأمر يختلف في نظم استرجاع الصور العلمية التي يمكن استخدام الصور ذاتها في عرض التخصصات الفرعية لموضوع نظام الاسترجاع العام. إلى جانب الاهتمام بوضع كل آليات بحث الصور في واجهة واحدة دون اللجوء إلى واجهات أكثر تقدما مما ييسر على المستفيد متابعة البحث. على أن تكون مختلف آليات البحث واضحة للمستفيد يختار منها باستخدام الترميز. ويعرض الجدول رقم (٤-٨) تفصيلا لعناصر واجهة البحث وعدد المحركات التي تبنت تلك العناصر، حيث يتضمن في شقه الأول العناصر المكونة لواجهات محركات البحث، ويركز الشق الثاني على مدى أهمية كل عنصر منها على حدة، إلى جانب عدد المحركات التي تختص بكل عنصر من مجموع عدد محركات عينة الدراسة.

الجدول رقم (٤-٨) يوضح جوانب واجهات محركات بحث الصور

جوانب الواجهات	عدد محركات	الأهمية في واجهة البحث
عرض بعض الصور	٣	غير مهم
واجهات بسيطة	٦	مهم
تعدد واجهات البحث	٤	مهم
شكل آليات البحث رموز وقوائم	قوائم ٥	الرموز تعلو القوائم المنسدلة في الأهمية
	رموز ٣	
المساعدة داخل واجهة البحث	٢	مهم

٥/٢/٤ طرق عرض النتائج ومكوناتها

إن اختصاص محركات بحث الصور بالتعامل مع وسيط ذا خاصية محددة مثل الصور، يعمل على النظر إلى الجوانب المساعدة التي يوفرها محرك بحث الصور والتي تساعد المستفيدين في تحديد وتقييم النتائج المسترجعة من قاعدة بياناته الداخلية. وتتمثل هذه الجوانب في عاملين اثنين أولهما شكل عرض النتائج المسترجعة من وحدات الصور؛ وهو ما يكون في استدعاء الصورة منفردة أو في مجموعة من الحقول الواصفة أو في مجموعة من البيانات الشارحة أو في مجموعة من البيانات المحددة للاستخدام. أما ثانيهما فهو مكونات الوحدة المسترجعة في كل محرك حيث تدل البيانات والحقول المصاحبة على طبيعة وخدمات محرك البحث التي يتيحها للمستفيدين. ويمكن التعرف إلى كلا العاملين في العناصر التالية :-

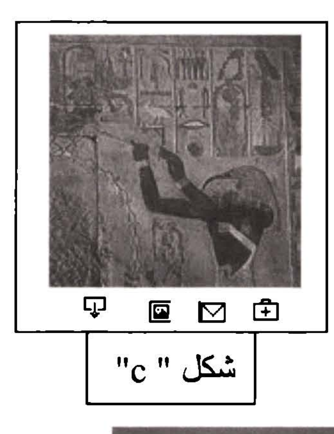
١/٥/٢/٤ طرق عرض النتائج من الصور

تعمل محركات بحث الصور على التميز فيما بينها في مختلف العناصر المساعدة على بحث الصور، والوصول إلى الدرجة المطلوبة من كفاية الاسترجاع،سواء كان ذلك في البحث أو عرض الصور، وشكل عرض النتائج هو أهم عناصر التباري بين محركات البحث. وتختلف أشكال عرض النتائج من لقطات منفردة أو لقطات مصاحبة لمجموعة من البيانات. وهي في محركات بحث العينة كما يلي:-



Pyramids_Egypt.jpg
282x200 pixels 13 KB
More info

شكل " a"



شكل " c"



شكل "d"

URL: http://www.cruisecyprus.com/paradise_cruises_israel_egypt.htm
Title: cruises to israel egypt holyland from cyprus
Filename:
http://www.cruisecyprus.com/paradise_cruises_israel_egypt.htm
File Size: 15 KB, 177 x 119

الشكل رقم (٤-١) لقطات نتائج محركات البحث

الجدول رقم (٤-٩) يوضح طرق عرض نتائج وحدات الصور الرقمية

	محركات بحث الويب						محركات بحث محددة	
	Alta-Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
صور منفردة	--	✓	--	--	--	--	✓	--
صور مع بيانات	✓	✓	✓	✓	✓	✓	--	✓
عدد الصور في واجهة العرض	١٥	٦	١٠	٢٠	١٦	١٥	٢٥	١٠
تحديد عدد الصور	x	x	x	x	x	x	✓	x

ويتضح من الجدول رقم (٤-٩) أن محركات البحث اعتمدت جميعًا على خيار عرض اللقطات الممثلة للصورة الحقيقية في مصاحبة مجموعة من البيانات كثيرة أو قليلة، ويرى الباحث أن خيار عرض لقطات الصور الرقمية يعد الشكل المناسب والملائم للتعامل المرئي مع الصور، وقد جاء الاعتماد على البيانات لمحركات العينة في أنها لا تعتمد في غالبيتها على تسجيلة ببليوجرافية يمكن الرجوع إليها للتعرف إلى مجموعة الحقول والعناصر الواصفة للصور، ومن ثم كان الاعتماد على البيانات. وقد استخدم محرك البحث Corbis شكلا مختلفًا بالإضافة إلى الشكل رقم (٤-١) وهو عرض اللقطات مع رموز أو رسومات Icons توضح الخيارات التي يمكن التعامل بها مع الصورة، كما يظهر في الشكل رقم (٤-١) شكل (C).

وهذا الخيار الذي يقترحه الباحث في ظل وجود تسجيلة ببليوجرافية، توضح كل أجزاء الصورة من موضوعات وشكل أمام المستفيد. وتكون اللقطات المفردة أكثر إيضاحًا إذا ما عرضت في غير وجود بيانات أخرى مما ييسر على المستفيدين استعراض النتائج.

وعلى الجانب الآخر، فإن عدد اللقطات أو الوحدات في واجهة العرض الواحدة له تأثير على التدقيق الذي يقوم به المستفيد في الاطلاع على نتائج البحث. وتتسم أربعة محركات من عينة محركات البحث بعرض عدد كبير جدا من الوحدات يصل إلى ٢٠ وحدة في واجهة البحث الواحدة، وهي Picsearch, Altavista, Visoo, Google. يترتب على أعداد الوحدات الكبيرة في العرض التشتت بين الصور وعدم الإجادة في الاطلاع، ويترتب على ذلك أيضًا صغر المقاسات للوحدات المرئية (اللقطات). وتتسم أربعة محركات أخرى هي Freefoto, Corbis Fast, Ditto بصغر الأعداد المرئية في واجهة العرض الواحدة إلى ست وحدات في محرك Ditto مما يعمل على التمكن الجيد الحكم على الوحدات المسترجعة. ويرتبط بهذا العنصر جانب آخر هو خيار تحديد الأعداد المسترجعة في كل واجهة عرض للنتائج وهو ما يمكن به للمستفيد التحكم في الوحدات المسترجعة إذا ما أراد التدقيق أو أراد التصفح السريع لكل النتائج، ويعطي محرك البحث Corbis فقط هذه السمة في البحث دون مختلف المحركات الأخرى.

٢/٥/٢/٤ أنواع البيانات المصاحبة لعرض النتائج

تختلف البيانات المصاحبة للصور المسترجعة في نتائج محركات بحث الصور بين بيانات الوصف المادي أو الموضوعي للصورة وبين تلك البيانات المتعلقة بطبيعة المحرك التجارية،أو العلمية، وخيارات التعامل مع كل وحدة من الصور المستدعاة. ويمثل الجدول رقم (٤-١٠) أنواع البيانات التي تصاحب نتائج استرجاع الصور وشكل البيانات من ألفاظ أو رسومات :-

الجدول رقم (٤-١٠) يوضح البيانات والحقول المناسبة لعرض نتائج الصور

		محركات بحث الويب						محركات بحث محددة	
		Alta-Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
حقول الوصف المادي للصور	عنوان الصورة	✓	✓	✓	✓	--	--	ID	✓
	مسار الملف		×		✓	--	✓	--	--
	موقع الصفحة المصدرية	--	✓	--	✓	✓	✓	--	✓
	حجم الملف	✓	✓	✓	✓	✓	✓	--	--
	مقاس الصورة	✓	×	✓	✓	✓	✓	--	--

ويتبين من الجدول رقم (٤-١٠) أنواع البيانات والحقول التي تتبع عرض الصور في محركات بحث العينة، ونتج أن محركي بحث الصور Google, Ditto قد حملا أكبر عدد من الحقول الواصفة للصورة الرقمية في عرض النتائج ككل، ويعمل محرك البحث Google بذلك على إكمال النقص في الوصف الببليوجرافي وإغفال المحرك للتسجيلة الببليوجرافية كاملة؛حيث اعتمد على بديل آخر هو عرض الصفحة المصدرية كوصف كامل للموضوع والشكل. أما محرك البحث Ditto فقد اعتمد على الشكل المبسط للتسجيلة الببليوجرافية المصاحبة للنتائج بدلا من إفراد الصفحات النهائية لمجموعات الصور. وتعتبر الدراسة محركي البحث Altavista, Fast من أهم محركات البحث التي اهتمت بعمل تسجيلة ببليوجرافية للصور الرقمية، وعلى الجانب الآخر فقد اهتمت هذه المحركات بعرض بعض البيانات عن الصورة في أثناء عرض النتائج ككل. وعلى الجانب الآخر فإن اختلاف محرك البحث

Corbis في الاعتماد على وضع خيارات التعامل مع الصورة دون إيضاح أية عناصر لوصف الصورة الموضوعي أو المادي، يرجع إلى طبيعة المؤسسة القائمة على هذا المحرك وهي الطبيعة التجارية، وهو ما يتضح جليًّا إذا ما تم النظر إلى الجدول الخاص بالتسجيلات النهائية لنجد الاهتمام بحقول النشر عن غيرها من الوصف الببليوجرافي للصور الرقمية. ويقترح الباحث تحويل البيانات إلى رموز تحدد كل ما يمكن التعامل به مع الصورة الرقمية، على أن يتم الرجوع إلى التسجيلة الببليوجرافية. ويفتقر محرك البحث Freefoto إلى البيانات التي تتبع الصور، على أنه في الوقت ذاته لا يحمل في تسجيلته البيانات الببليوجرافية الواصفة بالقدر المناسب مما يعكس مدى ضعف الاهتمام بمجموعات الصور داخل هذا المحرك، ويعزى ذلك إلى صغر حجم المجموعات في قاعدة بيانات هذا المحرك كما اتضح من أعداد النتائج.

٣/٥/٢/٤ سمات عرض الصور

انفرد محرك بحث Corbis بمجموعة من السمات الخاصة لاسترجاع الصور مثل إمكانية حصر البحث في الصور الملونة فقط، أو الأبيض والأسود فقط، وكذلك إمكانية اختيار الصور ذات الشكل الأفقي أو العمودي، وأيضا تحديد الصور التي يظهر فيها أشخاص والتي لا يظهر فيها أشخاص، وكذلك زاوية الصورة مثل صورة مباشرة أو علوية أو من الفضاء، هذا بالإضافة إلى إمكانية استدعاء صور منفردة أو ملفات صور. كما في الجدول رقم (٤-١١)

الجدول رقم (٤-١١) يوضح سمات عرض الصور

	محركات بحث الويب						محركات بحث محددة	
	Alta-Vista	Ditto	Fast	Google	Picsearch	Visoo	Corbis	Free-foto
تحديد لون الصورة (ملونة/أبيض وأسود)	x	x	x	x	x	x	✓	x
تحديد الشكل الهندسي للصورة أفقي/ عمودي)	x	x	x	x	x	x	✓	x

ونخلص مما سبق إلى أن شكل عرض الصور الرقمية في واجهة عرض النتائج يجب أن يكون خالصا من أي بيانات يمكن أن تعوق الاستطلاع للصور المسترجعة، إضافة إلى عرض أكبر قدر من مجموعات الصور على كل واجهة عرض، كما أن خلفية عرض النتائج يجب أن تكون باللون الأبيض الذي يوضح العناصر الداخلة فيه. هذا إلى جانب إضافة رموز تحدد الخيارات الممكن إجراؤها تجاه الصورة المختارة؛ من تنزيل أو شراء أو إرسال أو رابط للصورة أو الرابط إلى التسجيلة النهائية لعرض الصورة بجانب كامل بيانات الوصف الموضوعي والمادي.

٣/٤ خدمات محركات بحث الصور

تقدم محركات بحث الصور على الويب خدمات أخرى خلاف التي تبنى على بحث الصور ومن هذه الخدمات ما يلي:-

١/٣/٤ رسائل المتابعة أثناء البحث

رسائل المتابعة هي إحدى أدوات نظم الاسترجاع في التفاعل مع المستفيدين، وتبرز أهمية رسائل المتابعة في وضع المستفيد دائما على الطريقة الصحيحة لتحقيق الاستفادة الكاملة من مصادر معلومات هذا النظام. وتتركز رسائل المتابعة في شكل تصحيح آليات بحث قاعدة البيانات من خلال التعليق على كلمات البحث المدخلة إلى النظام. ولا تقف مثل هذه الرسائل فقط عند الإفادة بخطأ الكلمات المستخدمة، وإنما تعمل على إعطاء الشكل المناسب للبحث داخل قاعدة البيانات.

وقد اعتمدت بعض محركات بحث الصور على التفاعل مع المستفيد بواسطة رسائل المساعدة مثل Corbis، واعتمد متن الرسالة على عرض الآليات الصحيحة في البحث داخل هذه القاعدة. وعلى الجانب الآخر فإن محركات مثل Google عملت على تنبيه المستفيد بعدم صلاحية كلمات البحث دون الإرشاد إلى الشكل الصحيح لبحث قاعدة الصور. ويقترح الباحث أن تكون رسائل المساعدة هي ذاتها نسخة من صفحة المساعدة التي يوفرها محرك البحث لمستخدمي قاعدة الصور، يتم فيها شرح آليات بحث الصور.

٢/٣/٤ ترشيح الصور وفق سمات المستفيد

تعتمد بعض نظم الاسترجاع الحديثة على إستراتيجية بحث يمكن من خلالها إشراك المستفيد في استرجاع قدر أكبر من وحدات التحقيق، وهي البحث بالنتائج الأولية؛ حيث يطرح المستفيد كلمات البحث على أن يختار من النتائج المستدعاة اسم الموضوع أو رقم التصنيف المعبر عن الوجه الدقيق للموضوع. ومن ثم يمكن تضييق البحث من خلال إخراج مجموعة النتائج الأولية. غير أن هذه الإستراتيجية لا يتم العمل بها حرفيا في محركات بحث صور الويب، وقد اقترح محرك Freefoto من ذلك بترشيح أكثر الصور صلة بموضوع البحث أثناء تصفح المستفيد لوحدات النتائج.

٣/٣/٤ مرشحات الصور (الفلاتر)

تعد مرشحات الويب أو الفلاتر من أهم أدوات التعامل مع الشبكة العالمية التي تقدمها محركات البحث، وتمنع مرشحات صور الويب المواد غير المناسبة للاسترجاع في بيئة محددة. وقد عملت مختلف محركات البحث على تقديم مثل هذه الخدمة في بحث كل من النص والصورة من خلال إعداد قائمة توقف عن مجموعة المصطلحات التي إذا ما تضمنها مصدر المعلومات لم يمر من مصفاة محرك البحث. ويقترح الباحث أن يتم الاعتماد على قائمة توقف تحدد اتجاهات برنامج الزاحف في التعرف إلى أوعية المعلومات النصية والمرئية غير المناسبة، وهو ما يعد بمثابة سياسة التزويد والإضافة التي يقوم بها برنامج التزويد في محرك بحث الصور. وقد اهتمت محركات البحث Alta- Vista, Corbis, Fast, Visoo بالعمل على وضع آليات الترشيح للصور، للتعامل مع البيئات المختلفة لمستخدمي محركات بحث الصور الرقمية.

وعلى الجانب الآخر فإن لمرشحات الصور الأثر الأكبر في بحث واسترجاع محركات الصور؛ حيث تمنع المرشحات الكثير من وحدات الصور الداخلة في الاستدعاء أو الواقعة ضمن التحقيق في حالة تلبية الحاجة الموضوعية للمستفيد.

٤/٣/٤ كثافة الصور

تحتاج محركات بحث المعلومات المرئية على الويب إلى خصائص ترتبط بعرض الشكل المادي لوسيط المعلومات، ويرتبط بجودة الصور داخل محركات

البحث تقنية اختيار الصور ذات الكثافة العالية للنقاط داخل وحدة القياس الواحدة (البوصة)، فكلما زادت كثافة النقط زادت جودة الصورة في العرض. ويطلق على هذا المفهوم مصطلح Resolution. وقد أضاف محرك البحث Fast في تعريفه لهذه التقنية أن الصور الفوتوغرافية ذات الكثافة العالية من شأنها أن تأخذ حجما أكبر من الصور ذات الكثافة المنخفضة، كما أن عملية الطباعة لهذه الصور تتم بصورة أفضل من الصور منخفضة الكثافة، ويدخل في ذلك أيضًا حجم شاشة العرض أمام المستفيد؛ فتقل كثافة نقاط التوزيع بكبر حجم الشاشة والعكس، مما يؤثر على عرض الصور الرقمية[(١)]. وأشارت دراسة حول الصور الرقمية مدى الاهتمام بهذه التقنية سواء للصور التي نتجت عن كاميرات رقمية أو تلك التي نتجت عن الإدخال بالماسح الضوئي، حيث تؤثر الكثافة المنخفضة للصور الرقمية على إظهار الصورة داخل شاشات العرض أو أوراق الطباعة على أنها فقط مجموعة من النقاط المتراصة التي تمثل لوحة غير محددة المعالم.[(٢)]

ويتيح محرك بحث Corbis إمكانية الاسترجاع بدرجات مختلفة من الكثافة تتدرج من كثافة فوق عالية، وكثافة عالية، وكثافة متوسطة، وكثافة ضعيفة. ولم يذكر خيار الكثافة بين عناصر بحث أي من محركات بحث الصور جميعًا، سوى ما ذكره CORBIS في إرشادات البحث الخاصة به.

٥/٣/٤ حفظ سمات البحث للمستفيد لمرات قادمة

تعمل بعض محركات بحث الصور على تسجيل سمات البحث للمستفيدين من مجموعات الصور Profile. ومن ثم يمكن التعامل مع إستراتيجيات بحث ثابتة يستدعي بها المستفيدون ما يحتاجون إليه من الصور الرقمية، دون تكرارها بعدد مرات البحث.

وعلى الجانب الآخر، فإن محركات بحث الصور تعمل على تحليل هذه السمات لدراسة خصائص الصور المطلوبة إلى جانب الاحتياجات الموضوعية من

---

(١) Frequently Asked Questions - Picture Search. Fast Search & Transfer ASA, 2002, visited 20/12/2002,: http://www.victoria.tc.ca/Resources/www.services.html

(٢) Reichmann, Michael H. Understanding resolution, luminous landscape, 2002, visited20/12/2002,: http://www.luminouslandscape.com/understanding_resolution.htm cite

الصور مما يفيد أيضًا في تقنين خدمات الإضافة والتزويد لبرنامج الزاحف. ويعمل محرك Fast فقط على تقديم هذه الخدمة لباحثي مجموعات الصور داخل قاعدة محرك Alltheweb.

٤/٤ كفاية استرجاع محركات بحث الصور

إن الحديث عن معاملات قياس كفاية نظم وخدمات استرجاع المعلومات تأخذ أكثر من جانب يمكن تناوله، فقد تحدث لانكستر عن معاملات قياس كفاية خدمات المعلومات في مستويات ثلاثة؛ جاء في أولها الاعتماد على تقييم المستفيد لما يواجهه في التعامل مع نظام المعلومات وهو عن جوانب الزمن والتكلفة ثم العوامل النوعية في كفاية المحرك مثل جانب الاستدعاء وجانب التحقيق. وأخيرًا مدى الفائدة من المعلومات[(١)]؛ إلا أن الباحث يرى أن المعاملات المستخدمة لقياس كفاية خدمات المعلومات عند لانكستر لا تتطابق مع تلك التي يمكن الاعتماد عليها في قياس كفاية محركات بحث الصور، ويمكن إجمال حجة الدراسة في العناصر التالية مختتمة الحديث بنموذج آخر يلائم طبيعة عمل وتكوين محركات بحث الصور.

١/٤/٤ محركات البحث المحددة

لقد عمد الباحث إلى ذكر هذا النوع في بداية الدراسة على غير السابق من الخصائص؛ وذلك يرجع إلى التطابق البعيد بين نظام قياس الكفاية للانكستر والتعامل في محركات البحث المحددة التي تلائم مجموعات الملفات القريبة من الثبات في أعدادها. وتعتمد محركات بحث الصور وغيرها من قواعد بيانات خدمات المعلومات الأخرى على إحدى طرق التنظيم التي يبنى عليها شكل الاسترجاع مثل اعتماد الكلمات المفتاحية والاسترجاع بأسلوب التصفح، وتختلف محركات البحث في طبيعة التنظيم لمحتوياتها عن قواعد البيانات عامة، حيث الاعتماد على التكشيف للنص الكامل واعتبار كل كلمات النص نقاط استرجاع لوحدة المعلومات. إلى جانب ذلك، فإن وسيط الصورة يختلف في احتياجاته المعلوماتية عن تلك التي توجد للنصوص. وقد بدأ لانكستر بالحديث عن مقياسي

(١) لانكستر، ولفرد. نظم استرجاع المعلومات / ترجمة حشمت قاسم_. القاهرة:دار غريب للطباعة والنشر والتوزيع، ١٩٨١. ص ١٦٠.

الزمن والتكلفة وهما يعتبران وجهي عملة استخدام الإنترنت، حيث يترجم الوقت المبذول في البحث إلى تكلفة من جانب الباحث أو تكلفة في الوقت، ولا تختلف هذه المقاييس بين جميع نظم وخدمات استرجاع المعلومات؛ إلا أن تقرير الكفاية يتوقف على الوحدات المسترجعة من جانب النظم لإفادة الباحثين؛ فإذا ما كانت هذه الوحدات تلبي الجانب الأكبر من احتياجات المستفيد أصبحت التكلفة في مستوى منخفض إلى جانب الوقت المبذول لاسترجاع تلك الوحدات، وتسمى الوحدات المسترجعة المطابقة لاحتياجات المستفيد بالتحقيق؛ أحد أهم عوامل الحكم على كفاية الاسترجاع. وبالتعرض إلى تفسير نظام استرجاع لانكستر للمعلومات يتضح أن مفهوم الاستدعاء الذي يعتمد عليه المستفيد في تحديد كفاية النظام يختلف عما يوضحه الجدول رقم (٤-١٢) لنظام معاملات الكفاية للانكستر؛ حيث ينظر المستفيد إلى أن الاستدعاء هو كل ما تم استرجاعه من داخل قاعدة البيانات ككل : أي كل المسترجع. على حين أن معامل الاستدعاء عند لانكستر يتم حسابه اعتمادًا على معرفة عدد الوحدات المسترجعة ذات الصلة بالموضوع (وهو ما لا يمكن إدراكه داخل محركات البحث سريعة التحديث)، وعلى ذلك فإن الاستدعاء عند لانكستر هو فقط مجموعة من كل المسترجع؛ إلا أنه يساوي كل المسترجع عمليا عند المستفيد. ويوضح الجدول رقم (٤-١٢) معاملات الكفاية عند لانكستر :-

الجدول رقم (٤-١٢)

معاملات قياس الكفاية نموذج لانكستر يلائم لمحرك محدد[١]

	صالح للموضوع أ + ج	غير صالح للموضوع ب + د
مسترجع أ + ب	أ استدعاء تحقيق	ب شوشرة
غير مسترجع ج + د	ج فاقد	د استبعاد

(١) المصدر السابق نفسه.

اعتمادًا على الجدول رقم (٤-١٢) تكون معاملات قياس الكفاية كما يلي :-

معامل الاستدعاء = الوحدات المسترجعة الصالحة / كل الصالح للموضوع= أ / أ+ج
معامل التحقيق = الوحدات المسترجعة الصالحة / كل المسترجع = أ / أ + ب
معامل الشوشرة = الوحدات المسترجعة غير الصالحة / كل المسترجع = ب / أ + ب
معمل الفاقد = الوحدات غير المسترجعة الصالحة / كل الصالح = ج / أ+ج
معامل الاستبعاد= الوحدات غير المسترجعة غير الصالحة / كل غير المتصل= د/ج+د

٢/٤/٤ محركات بحث الويب

إن اعتماد نوع محركات البحث المحددة على نظام لانكستر يختلف بدوره عما يمكن التعامل معه في محركات بحث الويب، وذلك إنما يرجع إلى الديناميكية التي تعمل بها محركات بحث الويب في إضافة آلاف المواقع يوميا إلى قواعدها، فضلا عن اعتماد أساليب مختلفة في التنظيم والبحث لصفحات الويب وملفات الصور. إلى جانب البعد التام عن التدخل البشري في تنظيم وبحث محركات بحث الويب. ويجب ملاحظة وجود الاستدعاء في النوع السابق كجانب مكمل للفاقد لتكملة كل المتصل بالموضوع الذي لم يسترجع، على أن الناحية العملية تحتم إطلاق مصطلح الاستدعاء على كل الوحدات المسترجعة لكي تلائم التطبيق العملي لحكم المستفيدين. ويطرح الباحث النموذج التالي الممثل في الجدول رقم (٤-١٣) للاستخدام في تحديد كفاية محركات بحث الويب للصور الرقمية خاصة والأنواع الأخرى من محركات البحث عامة، كما يلي :-

الجدول رقم (٤-١٣) معاملات قياس الكفاية المقترحة لمحركات بحث الصور

	صالح للموضوع أ + ج	غير صالح للموضوع ب + د
مسترجع (الاستدعاء) أ + ب	أ تحقيق	ب شوشرة
غير مسترجع د	د استبعاد	

وتكون المعاملات كما يلي :-

معامل الاستدعاء = كل المسترجع / كل ما يشمله النظام = أ + ب / أ + ب + د

معامل التحقيق = كل المسترجع الصالح / كل الاستدعاء (المسترجع) أ / أ + ب

معامل الشوشرة = كل المسترجع غير الصالح/ كل الاستدعاء (المسترجع) ب/أ +ب

معامل الاستبعاد =كل المستبعد / كل ما يشمله النظام = د / أ + ب + د

تم بناء النظام السابق على أسس هي :-

- إن من يصنع التنظيم داخل محركات البحث هو برنامج (نظام المعلومات) ومن غير المؤكد معرفة عدد الوحدات التي تلبي موضوع المستفيد.
- إن استخدام مصطلح الفقد مع الاسترجاع في النظم الآلية لا يوافق بيئة الويب التي تعمل على تكشيف كل كلمات النص التي تصبح نقاط استرجاع للنص، ومن ثم فإن التحكم في استدعاء النصوص يرجع إلى المستفيد المحدد للموضوع. واعتمادًا على عدم تحديد النظام لما يحويه من وحدات عن موضوع بعينه، فإن حساب الفقد أو الفاقد من الوحدات ليس له معنى.
- إن طبيعة النظم الآلية تبنى على استدعاء كل ما يطابق إستراتيجية البحث ومن ثم لا يعطي النظام من جانبه أي شوشرة، وإنما يحدد المستفيد تلك الشوشرة من وجهة نظر احتياجات خاصة له. حيث يسترجع النظام كل ما تستدعيه كلمات البحث من وحدات.
- إن حقيقة التعامل مع نظم المعلومات الآلية هي أن الاستدعاء هو كل المسترجع وأن التحقيق هو ما يطابق احتياج المستفيد من داخل الاستدعاء.

ويتعامل النموذج السابق مع طبيعة محركات البحث على الويب التي تتعامل مع الموضوعات في صفحات الويب كما قام عليه مصممو هذه الصفحات. ينتج عن تكشيف محركات البحث للنص الكامل أن تكون مختلف كلمات النص مداخل وكلمات مفتاحية يتم الاسترجاع من خلالها. وحري بالذكر أن المحركات التي سيتم تطبيق النموذج السابق عليها هي محركات البحث للغة العربية التي تعمل بتقنيات متقدمة خلاف ما يوجد عليه الواقع الآني لمحركات البحث؛ كأن يتم استدعاء الأشكال المختلفة لكلمة مفتاحية واحدة تعكس كل الجوانب الموضوعية مثل (مكتبات) التي تسترجع كتب، يكتب، كتاب، مكتبة من خلال البحث بآلية الاشتقاق.

الجانب الثالث والأخير في طرح النموذج السابق أن الصورة كوسيط معلومات يتم تحليله موضوعيا تحتاج إلى وصف موضوعي يختلف عما توجد عليه وسائط المعلومات النصية.

٤/٥ الخلاصة

عملت الدراسة في هذا الفصل على تناول الخصائص العامة و النوعية لمحركات بحث الصور الرقمية، للعمل على قياس جوانب القوة والضعف بين محركات الدراسة، والأساليب المتبعة في تنظيم وبحث ملفات الصور الرقمية داخل قواعد بياناتها.وقد قسمت الدراسة هذه الخصائص إلى عامة في التطبيق بين كل المحركات و بين نوعية يعتمدها فقط بعض المحركات؛ لتنظيم وبحث الصور الرقمية.إلى جانب قياس خدمات محركات الصور مثل البيانات التابعة لعرض لقطات الصور وتقنيات العمل مثل قياس كثافة الصور الرقمية.



الفصل الخامس

مرشحات الصور

- ٥ / ٠ التمهيد
- ٥ / ١ مفاهيم الترشيح
- ٥ / ٢ ترشيح المعلومات المصورة
- ٥ / ٣ أنماط ترشيح المعلومات المصورة
- ٥ / ٤ أنواع المرشحات داخل محركات البحث
- ٥ / ٥ ترشيح الصور الرقمية
- ٥ / ٦ ترشيح الصور داخل محركات بحث الصور
- ٥ / ٧ مرشحات الصور داخل محركات البحث
- ٥ / ٨ العلاقة بين الترشيح ومعاملات نظم الاسترجاع
- ٥ / ٩ الخاتمة

مرشحـــات الصــــور

٥/٠ التمهيد

ينطوي مصطلح الترشيح بذاته على الكثير من المفاهيم التي قد تقترب في أحيان وتبعد في أحيان كثيرة عن مجال تنظيم وبحث المعلومات الرقمية. كما أن مفهوم ترشيح المعلومات على الويب يتضمن بداخله الكثير من المفاهيم الأخرى التي يمثل كل منها اتجاهًا في تناول ترشيح المعلومات، وصولا إلى مفهوم ترشيح الصور الرقمية داخل محركات بحث الصور. ويعتبر ترشيح الصور الرقمية أحد الموضوعات الفرعية المهمة داخل عمل محركات بحث الصور، فإن ثمة موضوعات جانبية لا يمكن إهمالها في عمل محركات الصور مثل ؛ كثافة الصور الرقمية، إدراج صور الإعلانات التجارية، إلى جانب الموضوع الحالي المتعلق بترشيح أو منع وحدات محددة من الصور الرقمية لا يتم إدراجها ضمن نتائج البحث.

إن تناول ترشيح الصور الرقمية على الويب يأخذ الكثير من المسارات المحددة بطبيعة المستفيدين. فيمكن تناول ترشيح الصور من ناحية نظم استرجاع المعلومات على ضوء التحكم في عنصري الاستدعاء والتحقيق، أيضًا الترشيح من ناحية الملاءمة لبيئة الاستخدام في منع الصور المخالفة، إلى جانب التعامل مع الصورة كشعار أو إعلان تجاري تؤخذ تجاهها سبل لمنع إزعاج مستخدمي الويب. وهناك الكثير من الدراسات التي اهتمت بترشيح المعلومات ليس من ناحية المحتوى وإنما من ناحية القيمة الموضوعية؛ حيث تميزت الويب عن غيرها من مصادر المعلومات باختلاط الغث بالثمين من المعلومات داخل الموضوع الواحد[١].

---

(١) Jacob, Palme. Information Filtering, Department of Computer and Systems Sciences,Stockholm University/KTH, 1998, cited 10/8/2005, http://dsv.su.se/jpalme/select/information-filtering.pdf

ويهتم هذا الفصل بالحديث عن مرشحات الصور لما لها من تأثير في تنظيم واسترجاع المعلومات المصورة داخل محركات بحث الصور. حيث يشمل الحديث آلية العمل لهذه البرامج، وتأثيرها على الإضافة والتنظيم داخل قواعد محركات البحث.

١/٥ مفاهيم الترشيح

١/١/٥ ترشيح المعلومات

يرتبط مفهوم ترشيح المعلومات filtering بمفهوم آخر ملازم له هو استرجاع المعلومات information retrieval، ويمكن القول إن ترشيح المعلومات هو جزء تابع للتحكم في المعلومات المستدعاة أو المسترجعة. وقد ورد مفهوم ترشيح المعلومات بحسب السياق المعني بهدف الترشيح، ومن بين التعريفات تعريف Nicholas Belkin " إن ترشيح المعلومات يساعد المستفيدين على مراقبة وتنقية المعلومات، أي أنه بشكل أكثر تحديدًا يتعلق بالتحكم في طبيعة المعلومات، وموضوعاتها الواردة إلى المستفيدين "(١). ويميل هذا المفهوم إلى طبيعة التحكم في المعلومات التي تعد الغاية الأساسية لوضع مرشحات البحث، ويرى الباحث أنه يتعارض مع السياق الذي ورد فيه المفهوم ؛ حيث تحدث السياق عن علاقة الترشيح باسترجاع المعلومات ومحركات البحث، غير أن السياق لم يعن باختيار المفهوم المحدد.

أما قاموس علم المكتبات والمعلومات على الخط المباشر ODLIS، فقد عكس المفهوم الحقيقي لمصطلح الترشيح المستخدم على الويب، وهو " استخدام برامج خاصة التصميم لمنع مستخدم الحاسب الآلي أو الشبكة أو النظام من عرض محتوى محدد سلفا من خلال غلق هذا المحتوى، والمرشح أو الفلتر يستخدم أساسا لمنع الأطفال من عرض المحتويات المخالفة على الويب "(٢). ويلاحظ الباحث تأكيد هذا المفهوم على الترشيح الأخلاقي للمعلومات التي تعد الوظيفة الأساسية لبناء برامج الترشيح.

---

(١) Belkin, Nicholas. Technical, Business, and Legal Dimensions of Protecting Children from Pornography on the Internet, NATIONAL ACADEMY PRESS, 2002, cited 1/8/2005, http://books.nap.edu/html/protecting_children/na_statement.html

(٢) ODLIS: Online Dictionary for Library and Information Science, cited 15/8/2005 http://lu.com/odlis/odlis_f.cfm#filtering

٢/١/٥ المرشحات أو برامج الترشيح

يرتبط بمصطلح ترشيح المعلومات مصطلح آخر هو مرشح المعلومات أو الفلتر الذي يدعم تحقيق مستخدمي الشبكات أو الحاسبات الآلية. وقد وضع ODLIS أيضًا تعريفًا لمرشح المعلومات filter هو " أحد برامج الحاسبات الآلية الذي يسمح لبيانات محددة مسبقًا بالمرور إلى المستفيد، فعلى سبيل المثال يسمح هذا البرنامج بمرور عدد محدد من رسائل البريد الإلكتروني الوافدة من مواقع سابقة التحديد ". [١] ومن ثم فإن ترشيح المعلومات في المعنى العام له **يعني البرنامج أو الإجراءات أو السياسات التي يضعها مصمم مواقع الويب أو مستخدم الحاسب الآلي للتعامل مع محتويات محددة سلفًا من المعلومات المسترجعة أمام المستفيد.**

وتعد المفاهيم السابقة ممثلة نسبيًا لواقع استخدام الويب خاصة في بيئات محددة مثل البيئات العربية الإسلامية التي تتعامل مع الويب. غير أن كل ذلك له تأثير على طبيعة استدعاء المعلومات والتحقيق. أما الكثير من المستفيدين الذين تختلف سماتهم them profiles في الوصول إلى معلومات الويب، وخاصة وحدات المعلومات المصورة.

٢/٥ ترشيح المعلومات المصورة

١/٢/٥ ترشيح المعلومات على الويب

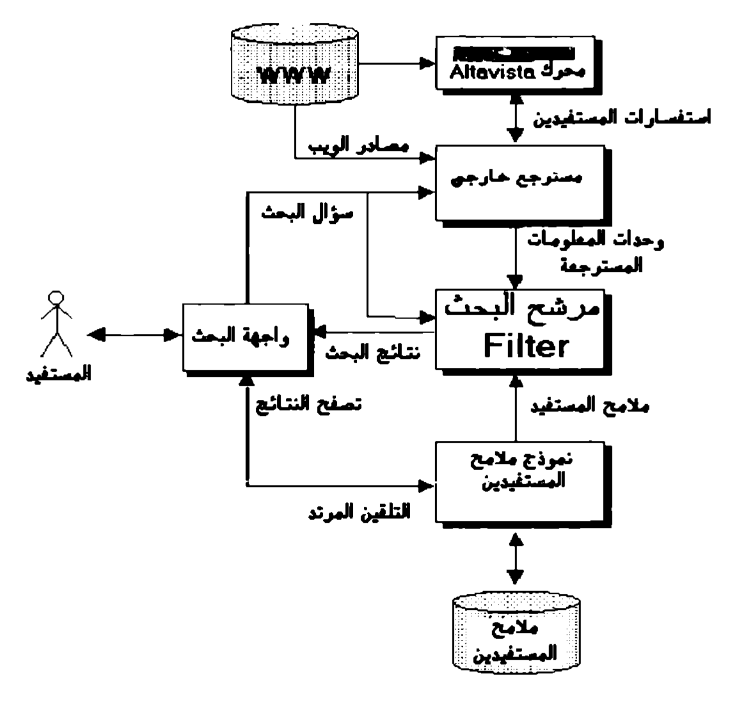
أضحت الويب المصدر الشامل لمصادر المعلومات على اختلاف موضوعاتها وأشكالها؛ النصية والصوتية والمصورة. وعملت الكثير من محركات بحث الويب على ملاءمة البيئات المختلفة لمستخدمي الويب بتوفير أدوات الحماية من المعلومات غير المرغوب بعرضها. وزاد أهمية ذلك وجود المواد المصورة الرقمية التي تعتبر أكثر مصادر المعلومات الرقمية حاجة إلى الترشيح والتنقية. مما أدى بمحركات بحث الصور خاصة إلى استخدام مرشحات الصور، وكان ذلك متمثلا في شقين هما ؛ الأول تصميم معايير ومقاييس إجراءات الترشيح، الثاني تصميم برامج

---

(١) Ibid.

ترشيح ذات واجهات بسيطة ليست رقابية؛ لتدعم المستخدم على تحديد المعلومات المهمة دون الغث من مصادر الويب.(١)

إن استخدام ترشيح الصور على الويب يعتمد على التحديد المسبق لمعايير استرجاع الصور داخل واجهة بحث المحرك. حيث ترسم كلمات الترشيح بالمرور أو التوقف stop/go الإستراتيجية العامة لاستدعاء الصور داخل قواعد بيانات الصور، ومن ثم تصبح الكلمات الدالة للترشيح هي كلمات ثابتة دائما في بناء إستراتيجية بحث الصور الرقمية داخل محرك بحث بعينه لمستفيد ذات سمات محددة. ويبنى على ذلك اختلاف معايير التحقيق والاستدعاء داخل محركات البحث التي تضع معها قياس مدى الكفاية في التوقف أو المرور لبعض وحدات الصور الرقمية داخل



الشكل رقم (٥- ١)

التفاعل بين سمات المستفيد وبحث مصادر الويب في ترشيح المعلومات

---

(١) Marinilli, Mauro. A Case-Based Approach to Adaptive Information Filtering for the WWW, Università di Roma Tre Via della Vasca Navale, 2002, cited 25/8/2005, http://wwwis.win.tue.nl/asum99/marinilli/marinilli.html

نتائج البحث. وقد مثل ذلك الحديث عن العلاقة بين صياغة نماذج ثابتة تعبر عن اتجاهات مستخدمي الويب في بحث صفحات الويب كما في النموذج التالي :-[١]

يبرز الشكل السابق تفاعل المستفيد مع أدوات بحث الويب في رسم ملامح بحث واسترجاع الصور الرقمية، وقد مثل محركات الصور في النموذج السابق محرك البحث AltaVista. يتم التفاعل بين المستفيد وواجهة بحث الصور على صياغة الكلمات الدالة على الموضوع مع الكلمات المفتاحية الدالة على الترشيح، ثم تهذيب وتحديد النتائج المستدعاة تبعًا للقياس الداخلي عند المستفيد في تحديد درجة التحقيق. كما يظهر الشكل رقم (٥-١) أن الترشيح في الصور الرقمية داخل محركات البحث يتم أثناء بحث المستفيد لقاعدة بيانات الصور الرقمية، وهذا يدل على أن الترشيح يأتي بحسب احتياجات كل مستفيد على حدة في استرجاع وحدات الصور الرقمية، يعكس ذلك عدم اتباع محركات البحث لسياسات الترشيح أثناء إضافة واقتناء وحدات الصور الرقمية باستخدام برنامج الزاحف. ويرى الباحث أن هذا يلائم فقط محركات البحث العالمية غير المرتبطة بثقافة ما مثل البيئة العربية أو الإسلامية. فإذا ما تقيد محرك بحث الصور بسياسة ثابتة لاقتناء وحدات الصور كان ذلك بمثابة ترشيح للصور الرقمية دون الحاجة إلى بناء قاعدة لسمات المستفيدين User Model.

يعتمد النظام السابق للتعامل مع المستفيد على رسم نموذج يحدد احتياجات المستفيد في بحث المعلومات المصورة. وهذا النموذج يبنى على محاور مثل :-[٢]

أولًا : المحتوى العام لعرض المعلومات الذي يلائم أو يختلف مع محتوى كلمات البحث لدى المستفيد المصاغ في كلمات البحث.

ثانيًا : مجموعة من كلمات البحث والترشيح حالية الاستخدام في بحث المستفيد، وهي بدورها تختلف حسب طبيعة كل مستفيد على حدة.

ثالثًا : مجموعة الأفكار أو الآراء المطروحة للبحث من جانب المستفيد، أو تلك المطروحة بشأن ترشيح المعلومات.

---

(١) Ibid.

(٢) Lbid.

رابعًا : مجموعة من الكلمات المفتاحية للترشيح التي تصاغ بعناية من قبل المستفيدين، على أن يكون لك من هذا الكلمات وزن نسبي يعكس طبيعة النتائج المسترجعة، بما يسمح بعد ذلك برسم إستراتيجية الترشيح لدى هذا المستفيد.

ويضيف الباحث محورًا آخر هو:

خامسًا : حاجة النظام إلى التحليل الدقيق لمحتوى المصادر والصور الرقمية على الويب، والتعبير عنها بمصطلحات كشفية أكثر دقة في وصف محتوى الصورة. وذلك لإحكام التفاعل بين محرك البحث ومستخدم وجهة البحث.

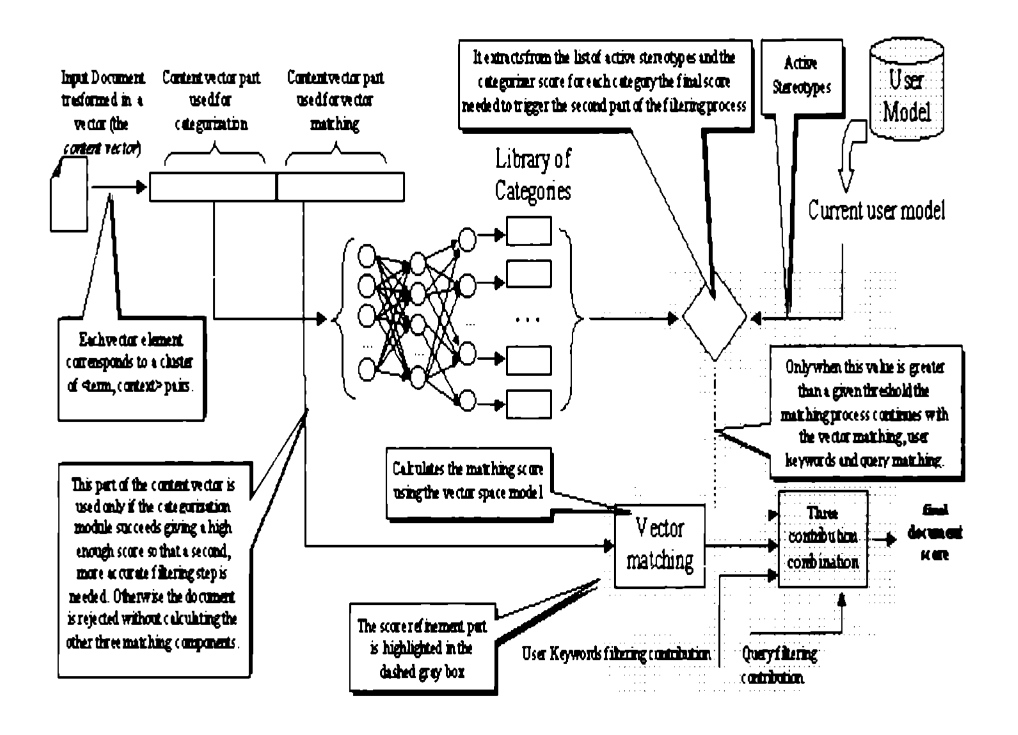
٢/٢/٥ آلية ترشيح المعلومات المصورة

تتم عملية ترشيح المعلومات بالتفاعل بين ما يقدمه محرك بحث الصور من مصادر، وما يحتاجه المستفيدون من الموضوعات المشار إليها بكلمات البحث. وعلى ذلك، فإن ثمة إجراءات تتم داخل محرك الصور تبدأ بانتقاء محرك البحث لوحدات الصور الرقمية باستخدام قوائم التوقف أو المرور في برامج الزاحف، ثم العمل على إظهار ملامح المحتوى الحقيقة داخل سياق الكلمات الكشفية عن الصور الرقمية المعالجة داخل برنامج المكشف على أن يتم تحديث سياسات التكشيف والإضافة في محرك البحث بما يستحدث من كلمات ترشيح داخل إستراتيجيات بحث مستخدمي الصور الرقمية. أيضًا فإنه يمكن الاعتماد على الكلمات المحددة سلفا من قبل المستفيدين في إجراء الإضافة والتكشيف.

وتتكامل العناصر التالية لوضع آلية ترشيح داخل محرك بحث الصور :-

- ✓ نموذج تحليل محتوى المصادر والصور الرقمية ثم التعبير عنها بقيم داخل ترشيح المعلومات.
- ✓ نموذج الاعتماد على كلمات الترشيح المسجلة من قبل المستفيدين التي تضاهي الكلمات الدالة على محتوى مجموعة الصور داخل قادة بيانات المكشف.
- ✓ نموذج استفسارات البحث المطابق للكلمات المفتاحية في محتوى الوثائق وكلمات الترشيح في نموذج المستفيدين.

ويشير الشكل رقم (٥-٢) إلى تفاعل النماذج السابقة لتفعيل آلية عمل الترشيح داخل محركات البحث.



الشكل رقم (٥-٢) آلية عمل الترشيح داخل محركات بحث الويب

يلاحظ الباحث في الشكل رقم (٥-٢) والنماذج السابقة مدى اتجاه محركات بحث الويب إلى صياغة سياسات الترشيح النصية، وهو ما يمكن اعتباره مركزًا على جانب واحد فقط من أنواع مصادر المعلومات على الويب. وتحتاج الصور الرقمية داخل محركات البحث إلى إضافة تحليل المحتوى بجانب تحليل النصوص. غير أن اعتماد محركات البحث على إضافة عناصر metadata يكسب سياسات الترشيح داخلها قدرة أكبر على ترشيح المعلومات المصورة.

٣/٥ أنماط ترشيح المعلومات المصورة

لا تختلف آليات الترشيح باختلاف هدف الترشيح سواء الأخلاقي أو الأمني أو تبعًا لقيمة المعلومات، وإنما تختلف أيضًا تبعًا لطبيعة المعلومات التي يتم

ترشيحها داخل الويب. ومن ثم يمكن الحديث هنا عن آليات ثلاث مختلفة في ترشيح المعلومات على الويب، ولكل من هذه الآليات ما يلائمها من خدمات وبرامج الترشيح على الويب.(١)

٥/٣/١ الترشيح داخل مصادر المعلومات

تتضمن الإنترنت الكثير من مصادر المعلومات التي تخالف ما يريده الآباء للعرض أمام أطفالهم، أو ما لا يريده اختصاصيو المعلومات داخل مراكز المعلومات المتصلة بالويب، هذا فضلا عن تقييم الباحثين لمجموعات المصادر المسترجعة داخل قواعد البيانات أو محركات البحث. وقد قدمت الويب برامج للتعامل مع هذا النمط من الترشيح منها :-

- ✓ Adblock Plus
- ✓ PithHelmet
- ✓ Privoxy
- ✓ AdSubtract

وتستطيع هذه البرامج بناء جدار مانع أمام الأشكال المختلفة للمعلومات غير المرغوب بها.

٥/٣/٢ الترشيح داخل البريد الإلكتروني

تستقبل الحاسبات الشخصية المتصلة بالويب الكثير من أشكال المعلومات الضارة في صورة رسائل البريد الإلكتروني. وتعمل بعض من برامج الترشيح التالية على وضع وصف ببليوجرافي لمجموعات الرسائل الإلكترونية مع انتقاء الرسائل المسموح لها بالمرور إلى المستفيد. ومنها :-

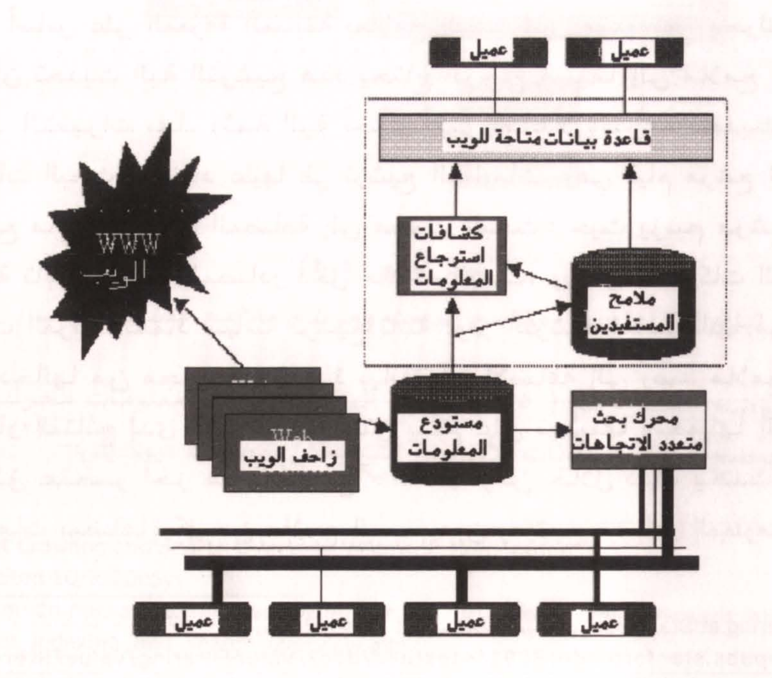
- ✓ Procmail
- ✓ maildrop
- ✓ Tagged Message Delivery Agent (TMDA)

---

(١) Gross, Ben. Information Filtering, 2005, cited 10/10/2005, http://bengross.com/filter.html

٥/٣/٣ الترشيح داخل محركات البحث

توفر محركات البحث مجموعة من آليات البحث ومعاملاته لتمكين المستفيد من صياغة إستراتيجية البحث النهائية داخل مجموعات قاعدة البيانات، وتتكون إستراتيجية البحث النهائية من الكلمات الدالة عن موضوع البحث، واعتمادًا على ذلك فإن كلمات الترشيح للانتقاء selection أو التوقف stop تؤثر بشدة على طبيعة النتائج المسترجعة من محرك البحث. ويرى الباحث أن استخدام المستفيدين لإستراتيجية بحث محددة وكلمات بعينها، إنما يعد في ذاته إستراتيجية لترشيح المعلومات المسترجعة من كامل مجموعات قاعدة البيانات. وعلى ذلك فان الكلمات الدالة على المحتوى يمكن مد استخدامها للتحول إلى كلمات دالة لترشيح المعلومات تلائم كل مستفيد على حدة، أو إستراتيجية ثابتة لترشيح المعلومات المضافة إلى قاعدة البيانات بصفة عامة. ويوضح الشكل رقم إشراك العنصر البشري لمراقبة وصياغة تدفق المعلومات إلى جانب الإسهام في رسم سياسات ترشيح المعلومات داخل محرك البحث. وهو ما أطلق عليه الترشيح المبني على قيمة المعلومات، التي يتم وزنها بالاعتماد على كلمات البحث والاستفسارات من جانب المستفيدين.



الشكل رقم (٥-٣) يوضح الترشيح المبني على القيمة داخل محرك البحث

يوضح الشكل رقم (٥-٣) أن ترشيح المعلومات داخل محركات بحث الويب يعتمد على طرح كلمات البحث لاسترجاع مصادر المعلومات المضافة بواسطة برنامج الزاحف. ويقوم مستودع المعلومات" قاعدة المحرك" بتلبية الحاجات الموضوعية للمستفيدين، ثم انتقاء وحفظ الكلمات البحثية داخل قعدة البيانات داخل قاعدة ملامح البحث feature repository. وهذه الملامح عبارة عن تقييم المستفيدين لطبيعة المعلومات داخل صفحات الويب. على أن بقوم المشرفون على محرك البحث بتهذيب تلك الملامح لتحويلها إلى سياسات مخصصة أو عامة لبحث قاعدة بيانات المحرك، غير أنه يمكن استخدامها يضًا كسياسات لإضافة مصادر الويب إلى قاعدة البيانات.[١] وعلى ضوء ما سبق تأخذ ملايين من صفحات الويب وزن أو درجة قيمة المعلومات داخل هذه الصفحات. وقد سعت الكثير من محركات البحث على رأسها Google في دعم ترشيح المعلومات حسب قيمة المحتوى.

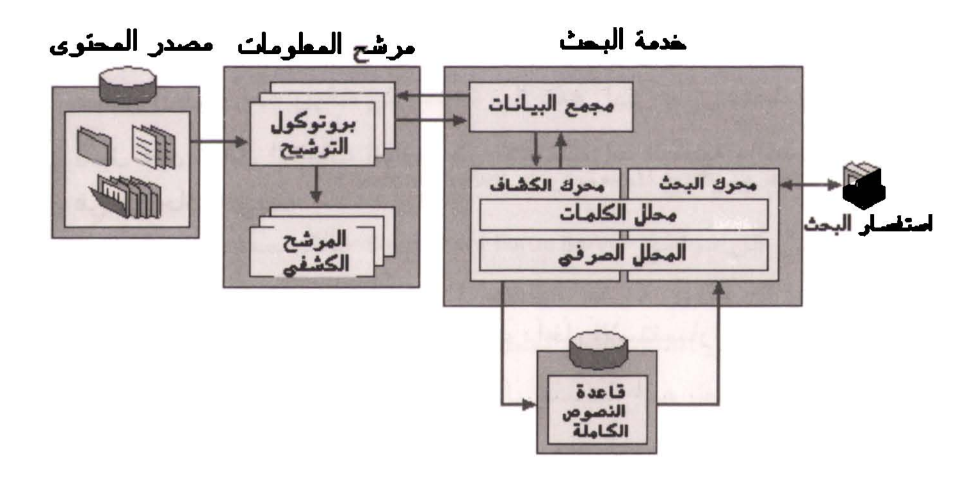
اعتمد النظام السابق في ترشيح المعلومات على مضاهاة ملامح البحث لدى المستفيدين بملامح الوثائق التي يتعامل معها محرك البحث. وهو ما يعتمد بشكل أساس على المعرفة المسبقة بملامح البحث لدى مستخدمي محرك البحث، غير أن تحديث آلية الترشيح هذه يحتاج الرجوع مسبقًا إلى ملامح الباحثين لرصد التغيرات بها. وثمة آلية أخرى أكثر تفاعلا وسرعة تحديث، تحاول محركات البحث الاعتماد عليها في ترشيح المعلومات، وهي قيام مرشح المعلومات بتجميع ملامح الوثائق المضافة إلى محرك البحث ؛ حيث يرسم مرشح البحث سياسة ثابتة لإضافة المصادر دخل محرك البحث. ويمكن لمحركات البحث في البيئات العربية اعتماد سياسة ترشيح ثابتة توفر الترشيح لمجموعاتها من الصور قبل إدخالها من مجموعات قاعدة بياناتها، بالإضافة إلى رصد ملامح البحث واختيار النتائج لدى المستفيدين التي يقوم على تهذيبها وتحويلها إلى محلل الوثائق عنصر آخر هو المُجمع gatherer، ومن خلال ذلك يحدث ترشيح المعلومات بمضاهاة كلٍ من ملامح المستفيدين وملامح مصادر المعلومات داخل

---

(١) Value Fltering at Stanford. [W.P.], 2002, cited 10/11/2005, http:/,dbpubs.stanford.edu:8091/~testbed/doc2/ValueFiltering/valueFilter-AdBW.ltm

محرك البحث.[1] كما توجد أيضًا الكثير من الأدوات المساعدة في تحليل الكلمات البحثية داخل مرشح محرك البحث مثل ؛ المحلل المعجمي word-breaker، المحلل الصرفي stemmers.

ويعد مرشح Share-Point Portal Server أفضل برامج الترشيح التي تستخدم هذه الآلية في ترشيح مصادر الويب داخل محركات البحث ؛ حيث يمكنها قراءة المعلومات الإضافية المختزنة داخل صفحات الويب metadata. وأيضا التعامل مع خادمات الويب وقواعد البيانات ذات المجموعات الكبيرة من مصادر الويب. ويوضح الشكل رقم (٥-٤) آلية التفاعل بين ملامح المستفيدين وملامح مصادر الويب داخل محرك البحث:-



الشكل رقم (٥-٤)

لبناء التفاعل بين ملامح البحث ومرشح ملامح مصادر الويب

يوضح الشكل رقم (٥-٤) البنية المكونة لمرشح معلومات محرك بحث الويب، ويوضح الشكل المكونات الأساسية التي يمكن تفصيلها فيما يلي:-[2]

---

(١) Content Crawling and Search Overview. Microsoft Corporation, 2005, cited 10/9/2005, http://msdn.microsoft.com/library/default.asp?url=/library/en-us/spssdk/html/_content_indexing_and_search_overview.asp

(٢) Lbid.

١) **محرك البحث**

يعمل هذا العنصر على تحويل الاستفسارات البحثية من اللغة الطبيعية إلى لغة SQL الخاصة بقاعدة البيانات.

٢) **محرك الكشاف**

يعمل هذا العنصر على تحليل خصائص صفحات الويب المأخوذة من نص صفحات الويب، ثم تحويل بعضها إلى قاعدة بيانات محرك البحث.

٣) **المجمع** gatherer

يعمل على تنظيم وإدارة مصادر الويب المضافة لمحرك البحث، إلى جانب احتوائه على القواعد المحددة للمعلومات حديثة الإضافة.

٤) **محلل الكلمات**

يعمل على تحليل الكلمات الواردة في الاستفسارات البحثية والنصوص المكشفة في صفحات الويب.

٥) **المحلل الصرفي Stemmers**

يعمل على التحليل الصرفي للكلمات داخل الاستفسارات البحثية ونصوص الويب.

٦) **مرشح المعلومات Filter Daemon**

وهو أهم عناصر ترشيح المعلومات، حيث يقوم باستيراد صفحات الويب حديثة الإضافة من المجمع، ثم تحويلها إلى سيل من المعلومات المرشحة "المفلترة" وخصائص تلك المعلومات محولا إياها إلى المجمع مرة أخرى. ويعتمد هذا العنصر على بروتوكولات تحليل النصوص إلى جانب قواعد الترشيح بالتكشيف IFilters.

٧) **بروتوكول الترشيح**

وهو عبارة عن مجموعة من البروتوكولات التي تجهز محتويات مصادر الويب للدخول إلى حيز الترشيح. وتلائم هذه البروتوكولات الأشكال المختلفة من مصادر المعلومات على الويب.

٨) **المرشحات الكشفية IFilters**

تقوم هذه المرشحات بفتح محتويات المصادر النصية ثم تفتيت نصها إلى خصائص وكلمات الترشيح.

٩) **مصادر المحتوى**

وهي مصادر الويب التي يتصفحها محرك البحث اعتمادًا على خصوصية محرك البحث في التعامل مع نوع بذاته من تلك المصادر. وتتحدد أنواع تلك المصادر بتحديد رابط الوصول إليها URL؛ حيث يبنى على ذلك أيضًا نوع البروتوكولات المستخدمة في ترشيح هذه المصادر.

٥/٣/٤ تسيير الترشيح داخل Share-Point Portal Server automation

يمكن تلخيص آلية عمل هذه المرشحات في شقين هما :-

٥/٣/٤/١ إضافة وترشيح المحتوى Content Crawling

وتعمل كل برامج Share-Point Portal Server بالاعتماد على زحف وجمع مصادر الويب ثم تكوين قواعد بياناتها للمعلومات المكشفة indices، إلى جانب احتواء المجمع على القياسات والمحددات التي تمكنه من التعامل مع ترشيح مصادر الويب ويحتفظ هذا العنصر بإحصاءات العمل النهائية لمرشح المعلومات. وتبدأ رحلة الترشيح تحديدًا عندما تعطي بعض ملفات التلميح بتغير محتوى صفحات الويب، وهي ملفات تحتفظ بمعلومات عن محتوى الويب داخل محركات ومرشحات الويب مثل ملفات NTFS، وتمد هذه الملفات عنصر المٌجمع برابط URL لبداية عملية التحديث والإضافة. وتتحول المصادر المضافة إلى محرك البحث من المٌجمع إلى مرشح المعلومات لاستقرائها في ضوء مواصفات الترشيح الموجودة لدى بروتوكول الترشيح لكي تتم عملية الترشيح بأفضل أشكالها، ثم تخرج كلمات الترشيح وبيانات مصادر الويب metadata بتفاعل كل من قواعد الترشيح النصي IFilters ومرشح المعلومات Filter Daemon. ثم تحول كل هذه البيانات إلى قاعدة بيانات محرك البحث لتحليل الكلمات المرشحة والدالة في مصادر الويب تحليلا لغويا وصرفيا للتعامل مع الاستفسارات البحثية القادمة من واجهة بحث المحرك.

٥/٣/٤/٢ تنفيذ استفسارات البحث Search Query Execution

يمرر محرك البحث استفسار المستفيد من واجهة البحث إلى المحلل اللغوي للتعامل مع كلمات التوقف والمرور في سؤال البحث. ثم تحول كلمات المرور إلى المحلل الصرفي لوضعها في الشكل المضاهي للكلمات الدالة داخل مصادر الويب ؛ ذلك كي تنتج عملية المطابقة والمقارنة بين الكلمات الدالة في الوثائق المفتاحية في الاستفسار. وبعد إتمام المضاهاة تدرج كل خصائص مصادر المعلومات المسترجعة للتأكد من موافقتها التامة لخصائص موضوع سؤال البحث، ثم تمر تلك المعلومات المسترجعة بوزن الأهمية في التحقيق، وينظم ثبت النتائج المستدعاة في شاشات عرض النتائج. ويحجب مرشح المعلومات وحدات المعلومات غير الملائمة لخصائص سؤال البحث.

٥/٤ أنواع المرشحات داخل محركات البحث

تختلف مرشحات البحث المستخدمة لمصادر معلومات الويب بحسب جانب الاهتمام بترشيح المعلومات. ويتجاذب محتوى الإنترنت أطراف عدة تبدأ بمصمم الصفحات أو منتج محتوى مواقع الويب، ومحركات البحث باعتباره وسيطًا لإيصال المعلومات إلى المستفيدين، ثم المستفيد النهائي الباحث عن المعلومات الرقمية. وباختلاف حجة لكل طرف من الأطراف السابقة يأتي وضع مرشح المعلومات كما في الأنواع التالية:-[١]

٥/٤/١ مرشح معلومات العميل Client-side filters

يثبت هذا النوع من المرشحات عند حاسب عميل البحث عن المعلومات الرقمية. حيث يمكن ضبط خصائص هذا النوع بما يوافق طبيعة البحث عن حاجات معلوماتية محددة. وتستخدم مرشحات العميل بجانب الحاسبات الشخصية، بالإضافة إلى الهيئات العاملة في بحث المعلومات على الويب كالمكتبات ومراكز المعلومات والشركات العاملة في الويب.

(١) FILTERING. National Academy of Sciences, 2003, cited 1/9/2005
http://www.nap.edu/netsafekids/pro_fm_filter.html

٥/٤/٢ مرشح معلومات مزودي الإنترنت Content-limited (or filtered) ISPs

يستخدم هذا النوع من المرشحات داخل مزودي خدمات الإنترنت لتحديد وصول مستخدمي الويب إلى نقاط موضوعية محددة. وتحتاط بذلك هذه الخدمات من استخدام بعض فئات المستفيدين مثل الأطفال لمواقع الويب غير الملائمة. فضلا عن قصر الاستخدام على بعض الخدمات التعليمية أو التسلية مثل مواقع المحادثات الكتابية أو الصوتية أو نشرات الأخبار الرقمية. ويعمل هذا النوع بكفاية للتحكم في إيصال رسائل البريد الإلكتروني المستقبلة من أفراد أو جهات محددة، بما لا يضر بالحاسبات المستقبلة الرسائل البريد الإلكتروني. ويلجأ مزودو الإنترنت إلى التحكم في المحتوى الرقمي بأشكال مختلفة ؛ كأن يحدد مزودو الإنترنت محتوى محددًا يصل إلى فئات الأطفال والبالغين من مستخدمي الويب. أما الشكل الآخر فهو طرح مستويات وصول إلى المحتوى تختلف بين الفئات العمرية المستخدمة لخدمات الإنترنت.

٥/٤/٣ مرشح معلومات خادمات الويب Server-side filters

يحقق هذا النوع أعلى كفاية في التعامل مع المعايير والضوابط الدولية لاسترجاع المعلومات؛ كنظم المكتبات و النظم الدراسية. حيث يمكن لأعضاء المؤسسة الواحدة الوصول إلى المعلومات بحدود ثابتة على نطاق أعضاء المؤسسة. ويمكن صياغة حدود الترشيح هنا بواسطة المؤسسات الدولية ذاتها أو خادماتها على الإنترنت.

٥/٤/٤ مرشح معلومات محركات البحث Search-engine filters

يستخدم هذا النوع داخل قواعد بيانات محركات بحث الويب. حيث يتيح محرك البحث إيقاف أو تشغيل مرشح المعلومات بحسب حاجة كل مستفيد على حدة. وتعمل هذه المرشحات عند تشغيلها إما على حجب النتائج غير المناسبة بين كل النتائج المسترجعة، ومنها Google, AltaVista، أو على تحديد نطاق البحث ضمن مواقع أو صفحات تلائم نوعًا محددًا من المستفيدين ومنها Yahoo, Lycos.

٥/٥ ترشيح الصور الرقمية

انحصرت أساليب استرجاع الصور الرقمية على الاسترجاع المبني على النص والاسترجاع المبني على المحتوى. ولعله يمكن القول إن هناك رابطًا ما بين

خصائص استدعاء وحدات الصور ضمن نتائج بحث محركات الصور وبين استدعاء مرشح الصور لمجموعة من هذه الوحدات ومن ثم حجبها عن العرض أمام المستفيدين. وبهذا المرتكز يمكن تفعيل آليات عمل مرشحات الصور في ترشيح المعلومات الرقمية المصورة. حيث يقارن مرشح الصور الرقمية بين الكلمات الواصفة للصورة وكلمات الترشيح لدى المستفيد. أما في الحالة الأخرى فإن مرشح الصور الرقمية يعمل على تحليل بنية الصورة الرقمية texture or content based مع نماذج البحث الصورة لحجب الصور المرشحة. وفيما يلي أنماط التعامل مع الصور الرقمية :-

٥/١/٥ الترشيح النصي

يعمل الترشيح النصي آلية التكشيف نفسها المستخدمة داخل محركات بحث الويب، ويعتمد الترشيح النصي على إخضاع النص المحيط بملف الصورة لاستخراج الكلمات الدالة على محتوى الصورة الرقمية، ومن ثم ترشيح محتوى الصورة اعتمادًا على كلمات النص الدالة. فضلا عما يمكن أن تفيد به عنصر ميتاداتا في تخصيص عنصر محدد داخل حقول وصف الصورة، يضم هذا العنصر جميع الكلمات الدالة التي تحويها الصورة، ويمكن في هذه الحالة نسخ محتويات عنصر الكلمات المفتاحية meta keyword أو تهيئة حقل الكلمات المفتاحية ذاته للتعامل مع مرشحات البحث. ويقترح الباحث هنا تصميم سياسة ثابتة بين عمل مرشح البحث وحقول ميتاداتا تمكن مرشح البحث من تحليل محتوى الصورة دون الاعتماد على النص الخارجي لصفحة الويب.

وتستخدم مرشحات الصور أساليب ثلاثة مختلفة للوصول إلى الترشيح المبني على النص؛ وهي كما يلي :-[١]

٥/١/١/٥ إيقاف الكلمات الدالة Keyword Blocking

يتخذ مرشح الصور هنا قائمة بمجموعة من الكلمات الدالة المحددة سلفًا التي إذا ما وجدت في ملف أو صفحة الصورة الرقمية توقف محرك البحث عن

---

(١) Davies, Roy. Vision and Signal Processing. London university, 2004, cited 10/10/2005, http://www.ph.rhbnc.ac.uk/research/sp/sp.html

ضمها إلى قاعدة بياناته. ومن ثم يغلق محرك البحث وصول المستفيدين إلى صفحة أو موقع الويب من خلاله؛ إلا أن هذا الأسلوب يواجه مشكلة أساسية في اعتماده على الترشيح بالنص فقط، وهي أن الترشيح لا يلقي أهمية إلى سياق النص الذي وردت فيه الكلمات الدالة، فعلى سبيل المثال يمكن لكلمة مثل (الجنس) أن تأتي في صفحة عن الأدوية أو الاستشارات الطبية لا تحمل أي صور تعبر عنها الكلمة، في حين يمتنع عن ضمها مرشح البحث.

٢/١/٥/٥ إيقاف المجموعات الموضوعية Packet Filtering

يعمل مرشح البحث هنا على غلق الوصول إلى نوع محدد من عناوين الإنترنت IP addresses الدال على طبيعة صفحاتها. وهذه الطريقة تعد سهلة الاستخدام داخل مرشحات البحث؛ إلا أن تطور تقنيات عناوين الإنترنت قد تخدع هذه المرشحات في إعطاء صفحاتها عناوين وهمية IP-independent virtual hosts تخالف الطبيعة الحقيقية لمحتوى موقع الويب.

٣/١/٥/٥ إيقاف محدد المصادر URL Blocking

يتعامل مرشح البحث في هذه الحالة مع كل موقع/صفحة ويب على حدة، خلاف الأسلوب السابق. وتبنى آلية العمل هنا على تجهيز قائمة بكل مسارات صفحات الويب مغلقة الوصول URLs. ويستطيع المستفيد هنا تصفح محتويات المواقع خلاف المسارات التي يمتنع مرشح البحث عن عرضها لمخالفة محتوى هذا المسار.

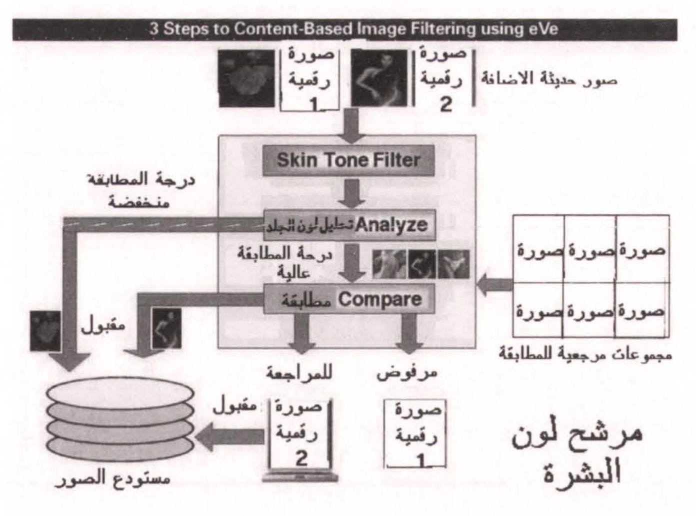
وعلى الرغم من تميز كل من الطرق السابقة في غلق المحتوى غير الشرعي أمام مستخدمي الويب، إلا أنها مازالت بحاجة أكبر للتطوير والتحديث، بل ويمكن القول أيضًا إن الويب بوصفه بيئة عالمية مفتوحة لمختلف الموضوعات ومختلف اللغات يصعب السيطرة على محتواها من خلال استخدام مرشح بحث ذي لغة واحدة أو مرشح يحمل كلمات دالة عن موضوع واحد.

٢/٥/٥ الترشيح التحليلي

إن استخدام الكلمات الدالة في الترشيح النصي يؤدي دوره بكفاية داخل وسيط المعلومات النصي، غير أن ذلك لا يحقق المرجو منه مع وسائط المعلومات

المرئية والمسموعة. وأخذ نمط الترشيح التحليلي الكثير من أشكال التعامل التقني الرقمي مع الصور على الويب. وقد عمل فريق أكاديمي من جامعة لندن على إخضاع دراسات معالجة الإشارات الرقمية للتطبيق على تحليل محتوى الصور الرقمية signal processing for digital image. وتعتمد هذه الدراسات على تحليل محتوى الصور الرقمية بما يمكن معه تحديد مختلف العناصر المرئية التي يبنى عليها محتوى الصورة.ومن ثم تستطيع مرشحات الصور المبنية على معالجة الإشارات المرئية تحويل ملامح الترشيح لدى المستفيدين إلى خصائص مرئية رقمية، يستخدمها مرشح المعلومات في مضاهاة وحدات الصور الرقمية المسترجعة داخل قواعد بيانات محرك الصور.[١]

وقد اهتمت معظم محركات بحث الصور الرقمية خاصة بالتعامل مع مرشحات صور البالغين ؛ حيث يعد هذا الجانب أكثر جوانب الحاجة إلى استخدام مرشحات الصور على الويب. تعتمد مرشحات الصور الرقمية الداعمة لتحليل المحتوى داخل هذا المجال على ثلاث مراحل يوضحها الشكل رقم (٥-٥) كما يلي :-[٢]



الشكل رقم (٥-٥) لمراحل الترشيح المبني على المحتوى للصور الرقمية

---

(١) Ibid.

(٢) using eve for Content Filtering. EVision, 2004, cited 5/10/2005, http://www.evisionglobal.com/business/eVe_content_filtering.pdf

٥/٥/٢/١ مرشح لون البشرة Skin Tone Filter

والهدف الأساس من هذه الخطوة معرفة المناطق الملونة ذات المساحات الكبيرة على سطح الصورة الرقمية. حيث خصصت هذه المرشحات تركيبة من الألوان الملائمة للتعرف إلى ألوان تشابه إلى حد كبير ألوان بشرة الإنسان. حيث اعتمدت على اختيار اللون الأحمر مع مزيج من البني والأصفر القاتم للتعرف إلى الأشكال المختلفة التي تتواجد عليها هذه الصور. وتبعًا لأن الألوان السابقة هي التي تشغل المساحات الكبيرة من بنية الصورة، فإن مرشحات البحث تفصل تلك الصور عن مجموعات الصور داخل قاعدة بيانات محرك البحث. ويمكن القول إن الألوان المختلفة ذات المساحات الضيقة تشير إلى عدم ظهور البشرة بشكل كبير في الصورة الرقمية، ومن ثم لا يمنع مرشح البحث إضافتها لمجموعات الصور.

وتعتمد بعض الفلاتر في تحليل ألوان الصورة الرقمية على طريقتين ؛ أولا طريقة قمة اللون height وأبرز صورها طريقة 256 color ؛ حيث يعطي مرشح الصور قيما ثابتة لكل لون ثم تحليل كل نقطة pixel على الصورة وتحديد قيمتها، ومن ثم حساب قيمة الصورة ككل. ثانيًا طريقة color based فتتعامل مع الألوان بحسب مكوناتها الأصلية red/green/blue. وهذه الطريقة تعتبر أسرع في قراءة ألوان الصورة والتعبير عنها من الطريقة السابقة.(١)

٥/٥/٢/٢ تحليل الصورة Analyze

تأخذ الصور المستبعدة من الإضافة إلى محرك البحث التوقيع الرقمي Automatic segmentation الذي تعرف من خلاله ملاءمة أو مخالفة الصورة عن سياسة الترشيح داخل محرك البحث.

٥/٥/٢/٣ المضاهاة والمقارنة comparison

بعد تحديد التوقيع الرقمي لمجموعات الصور، توضع الصورة الرقمية موضع المضاهاة بينها وبين مجموعات مرجعية سالفة الإضافة داخل قاعدة بيانات محرك البحث، وهذه المجموعات تحمل مجموعة من الخصائص يستطيع مرشح

---

(١) Image Filtering. No software patents, 2003, cited 1/10/2005, http://www.nosoftwarepatents.com/

الصور تحديد القرار النهائي باستبعاد أو إضافة الصور الرقمية موضع الترشيح. ويتم إجراء المضاهاة بأن تستبعد تماما الصور التي تحصل على ٧٠٪ فأكثر لنسبة التشابه بينها وبين مجموعات الصور المرجعية. أما الصور ذات نسبة ٤٠-٧٠٪ فيتم إرجاؤها لتدخل العنصر البشري في قرار الاستبعاد أو الإبقاء. وتضاف إلى قاعدة بيانات محرك البحث مجموعة الصور التي حصلت في مضاهاتها على نسبة أقل من ٤٠٪.

وقد اتفقت هذه الأنواع داخل مرشحات البحث مع تحليل محتوى لأغراض الاسترجاع، وتبني عناصر تحليل محتوى الصورة الرقمية عامة على ما يلي :-[١]

أولًا: الألوان الكونية Global Color

يعتمد تحليل المحتوى هنا على تحديد القيم الثابتة والمتغيرة للألوان داخل الصورة الرقمية اعتمادًا على تحليلها إلى الألوان الكونية الثلاثة الأحمر والأخضر والأزرق، ومن ثم تأخذ الصورة كاملة قيمة واحدة.

ثانيًا: الألوان المحلية Local Color

يعمل هذا العنصر على تحديد اللون الغالب داخل الصورة الرقمية، ثم إعادة تحليل هذا اللون اعتمادًا على قيمة واحدة بين ٢٥٦ قيمة لكل الألوان المستخدمة في تحليل الصور الرقمية.

ثالثًا: شكل الصورة Shape

يستخدم هذا العنصر تقنيات تحليل الصورة إلى الأشكال الرقمية المكونة لمحتوى الصورة، واعتمادًا على ذلك تتضح خصائص كل شكل على حدة بما يمكن التحكم في جميع أشكال الصورة.

رابعًا: بنية الصورة Texture

يتلخص هذا العنصر في تحليل اتجاهات بنية الصورة ودرجة التباين داخلها.

---

(١) Venters, Colin C. and Cooper, Matthew. A Review of Content-Based Image Retrieval Systems, University of Manchester, 2002, cited 1/10/2005, http://www.jtap.ac.uk/reports/htm/jtap-054.html

عملت أنواع المرشحات السابقة على ضبط نوع واحد من الموضوعات غير المشروعة على الويب، وقد اجتهدت مرشحات التحليل المرئي على تحليل بنية الصورة. غير أن الويب تتضمن من النصوص ما هو أخطر من الصور أيضا. وعلى ذلك يرى الباحث أن **استخدام تقنية تحليل المحتوى المرئي للصور الرقمية بجانب تحليل النص يصنع مرشح بحث متكامل قادر على ضبط محتوى الويب المرئي متعدد اللغات والموضوعات.**

٦/٥ ترشيح الصور داخل محركات بحث الصور

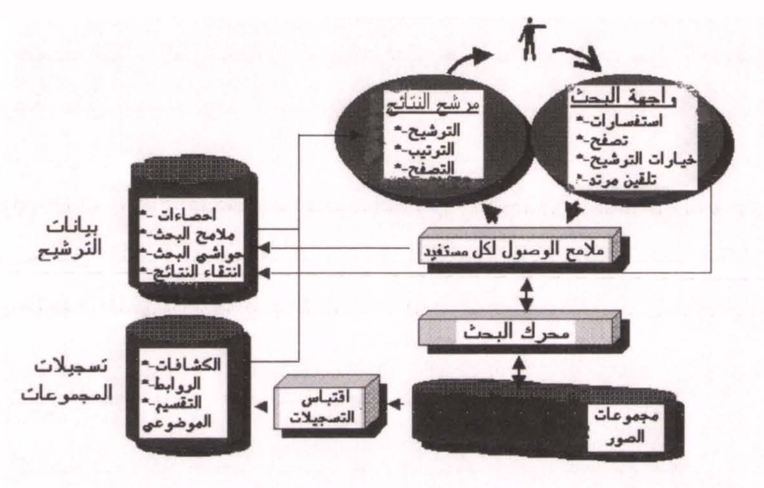
تنتهج محركات بحث الصور منهجًا خاصًا في التعامل مع الصور الرقمية؛ إضافة وتنظيمًا وبحثًا. وهي تختلف بذلك عن التعامل مع الأشكال الأخرى من المعلومات الرقمية. وتأخذ مرشحات الصور بعض خصائص التكشيف والاسترجاع داخل محرك الصور. فيأخذ من المكشف مسح النصوص المصاحبة للصورة أو مسح عناصر ميتاداتا ذاتها داخل ملف الصورة الرقمية، ومن ثم يمكنها التعرف إلى الكلمات الدالة على الموضوع ثم مضاهاتها بما لديها من قوائم كلمات التوقف. هذا فضلا عن إمكانية تزويد مرشحات الصور بالتقنيات اللازمة لتحليل بنية وألوان الصور الرقمية لاستخدامها في ترشيح الصور. وتكون بذلك الألوان في الترشيح التحليلي مرادفة لاستخدام الكلمات في الترشيح النصي.

ويهتم هذا الجزء من الدراسة بالحديث عن خصائص عمل مرشحات الصور داخل محركات الصور وتحديدها لإمكانية دراسة واقعها الفعلي داخل محركات البحث في الجزء التالي من الدراسة. ويعتمد الباحث في تحديد خصائص مرشحات الصور على النموذج الوارد في الشكل رقم (٦) :-[١]

يركز شكل رقم (٥-٦) لترشيح الصور الرقمية داخل محركات البحث على استخدام الترشيح النصي، ويتضح ذلك جليًا من تحليل الميتاداتا وتعليقات المستفيدين وتسجيل ملامح مستخدمي الصور الرقمية داخل محرك البحث. وقد استخلص الباحث الخصائص التالية التي يعتمد عليها مرشح الصور:-

---

(١) Value Filtering at Stanford. Optc.



الشكل رقم (٥-٦) بنية ترشيح الصور داخل محركات البحث

١. إمكانية تحليل نصوص صفحات و/ أو ملفات الصور الرقمية " الميتاداتا" لاقتباس الكلمات الدالة داخل محتوى الصورة الرقمية حالية الإضافة لقاعدة البيانات.

٢. إمكانية تحليل روابط بنية صفحات الويب حال اعتماد مرشح البحث على الترشيح باستخدام الروابط URL blocking.

٣. العمل بإحدى الطريقتين في الاعتماد على قوائم التوقف السابقة لإضافة الصور واعتمادًا على تحليل تعامل مستخدمي الويب مع تلك المجموعات سواء باستبعادها أم باختيارها، وهي إحصاءات تساهم في ترشيح مجموعات الصور.

٤. السماح بتسجيل ملامح المستفيدين الموضوعية المدخلة قبل مرحلة البحث.

٥. السماح بإتاحة مجموعات الصور لتحكيم العنصر البشري سواء أمام المستفيدين أم أمام مصممي مرشح الصور.

٦. إمكانية تحكم المستفيدين في تحديد مستوى ترشيح الصور قبل إجراء البحث، وزيادة خيارات التحكم في مرشح الصور حيث لا تقتصر فقط على غلق أو فتح مرشح الصور.

٧. استخدام تقنيات تحليل بنية الصور الرقمية لرفع كفاية ترشيح الصور؛ إضافة إلى تكامل هذه التقنية بتحليل محتوى الصورة بأبعاده ؛ البنية والشكل واللون.

تشكل الخصائص السابقة أساس بناء مرشح بحث للصور الرقمية على الويب، غير أن ضبط عمل هذه المرشحات يعتمد أولا على الاتجاهات التجارية لمحركات بحث الويب، فإن درجة اهتمام محرك البحث باستخدام المرشحات تنبني على الهدف العام لمجموعات الصور داخل قاعدة البيانات. وحري بالذكر هنا أنه لا مجال للمقارنة في ذلك بين قواعد بيانات الصور ومحركات البحث ؛ حيث إن الأولى تعتمد منذ البداية على العنصر الإنساني المدرك تماما لطبيعة محتوى مجموعات الصور الرقمية.

إن الخصائص السابقة قد صاغها الباحث اعتمادًا على الدور المهم الذي أداه مرشح البحث في رفع أو خفض مستوى الاستدعاء والتحقيق. غير أن هناك الكثير من الخصائص التقنية التي تدخل ضمن اختيار مرشحات الصور الجيدة، خاصة وأن هدف الترشيح لذاته هو الغاية الأهم في بعض البيئات مثل العربية والإسلامية، خلاف ما يترتب على ذلك من تأثر معاملات الاسترجاع. واشتق الباحث الخصائص التقنية لمرشح الصور الملائمة للصور الرقمية العربية، مما جاء عن مرشح content-protect وهو من أبرز مرشحات الصور المستخدمة في مجال الويب. كما يلي :-[١]

- ✓ يمتلك مرشح البحث إمكانية تحليل كلمات النصوص أو قوائم مسارات مواقع الويب، فضلا عن التفاعلية في تحديث وتغيير بيانات تلك القوائم.
- ✓ إمكانية ترشيح أكبر مجموعة من الصور بأقل وقت ممكن.
- ✓ ملاءمة الأشكال المختلفة من لغات صفحات الويب، والتعامل مع مختلف أشكال الصور الرقمية مثل PNG, GIF.....
- ✓ تصميم واجهة تطبيق سهلة الاستخدام أمام مستفيدي الويب/ خاصة غير الماهرين منهم.
- ✓ إمكانية التكامل مع جميع محركات البحث على اختلاف تخصصاتها الموضوعية، وأشكال المصادر الرقمية.

---

(١) Content Protect Web Filtering. Content watch, 2005, cited 1/9/2005, http://www.contentwatch.com/products/contentprotect.php

- ✓ إمكانية التحول التلقائي عند اختيار شكل البحث ؛ حيث يتحول مرشح البحث إلى الوضع الآمن عند تحديد المستفيد للبحث في شكل الصور فقط.
- ✓ إمكانية التحكم والتغيير والصيانة الدائمة لمرشح البحث في إعدادات العمل المختلفة؛ مثل طريقة الاتصال واختلاف مراحل البحث.
- ✓ إمكانية تحرير تقارير الحالة التي تعبر عن حالة الاستخدام وحالة الكفاية لمرشح البحث.
- ✓ إمكانية إدخال التعديلات من جانب المستخدم ؛ ذلك لضبط وتحديد قوائم التوقف بحسب حاجة المستفيد.
- ✓ إمكانية العمل برسائل المساعدة الداعمة للتفاعل مع المستفيدين أثناء ضبط خيارات البحث في مرشح الصور.
- ✓ إمكانية العمل بلغات مختلفة ؛ سواء لغات واجهات البحث أم لغات الترشيح لمصادر معلومات الويب.
- ✓ إمكانية إدخال التحديثات تلقائيا اعتمادًا على ملامح بحث المستفيدين ومصادر المعلومات موضع الترشيح.

إن الخصائص السابقة يمكن الاعتماد عليها في تشكيل مرشح الصور العربي الداعم لخصائص اللغة العربية، فضلا عن إمكانية التحكم في وضع هذا المرشح ضمن مراحل الإضافة والتزويد لمجموعات الصور أو عند إجراء البحث والاسترجاع من جانب المستفيد.

نهاية، يمكن صياغة الخصائص السابقة في محاور أساسية يمكن دراستها في واقع عمل محركات بحث الصور كما يلي:-

١) تفعيل مرشح الصور داخل محركات بحث الصور.

٢) مرشح الصور بين مراحل عمل محرك الصور.

٣) العلاقة بين مرشح الصور ومستخدم الصور الرقمية.

٤) نمط الترشيح المتبع داخل مرشح الصور.

٥) فاعلية مرشحات الصور داخل محركات البحث.

٥/٧ مرشحات الصور داخل محركات البحث

تختلف محركات بحث الصور في درجة اهتمامها بترشيح مجموعاتها من الصور الرقمية. وعلى إثر هذا الاختلاف تختلف درجات استخدام مرشحات البحث في عدم وجودها داخل بعض المحركات إلى وجودها دون خيارات الترشيح داخل البعض الآخر، ثم إلى التفاعلية مع المستفيد في تحديد خصائص وحدات الصور المراد استدعاؤها داخل قاعدة البيانات. وقد عملت أربعة فقط من محركات بحث عينة الدراسة على استخدام مرشحات البحث لمجموعاتها بدرجات متفاوتة. وهذه المحركات هي:-

١. AltaVista
٢. Corbis
٣. Fast
٤. Google

وسوف يتم دراسة واقع ترشيح الصور الرقمية داخل محركات البحث تبعًا للمحاور المذكورة، وهي كما يلي :-

٥/٧/١ تفعيل مرشحات الصور:-

تختلف فاعلية مرشحات الصور داخل محركات البحث السابقة في استرجاع الصور بحسب طبيعة خيارات البحث بها، وبالنظر إلى الجدول رقم (٥-١) يتضح مستوى فاعلية مرشحات الصور في هذه المحركات. كما يلي :-

الجدول رقم (٥-١) مرشحات الصور داخل محركات بحث العينة

محركات البحث		AltaVista	Corbis	Fast	Google
درجة الاستخدام	فتح / غلق	×	×	✓	×
	خيارات الترشيح	✓	✓	×	✓
خيارات البحث	الترشيح بحسب الشكل	✓	×	×	✓
	وضع كلمة مرور	✓	✓	×	×

تتسم مرشحات الصور داخل محركات بحث العينة بخصائص تفعيل كل محرك على حدة، وهي كما يلي :-

أولًا : أتاحت محركات بحث Google,AltaVista خيارات الترشيح على استحياء، حيث مكنت مستخدميها من تحديد الشكل موضوع الترشيح إذا ما كانت الصور الرقمية فقط كشكل محدد أم ترشيح صفحات الويب بجانبها.

ثانيًا : اقتصر محرك البحث Fast على طرح خاصية تنشيط مرشح الصور أو غلقه دون التطرق إلى خيارات ترشيح إضافية، ويأتي ذلك من طبيعة محرك Fast كمحرك متعدد البحث واعتماده على البحث في محركات بحث أخرى.

ثالثًا : **غلبت الطبيعة التجارية لمحرك البحث** Corbis على أداء مرشح البحث ؛ حيث يتيح Corbis خصائص مرشح البحث بعد التسجيل في قائمة مستخدمي Corbis. وحري بالذكر أن Corbis لا يستخدم مرشح البحث لحجب مجموعات الصور الخارجة، وإنما لتهذيب النتائج ورفع درجة التحقيق فقط داخل النتائج المسترجعة؛ أي أنه أداة من أدوات الاسترجاع فقط.

٥/٧/٢ العلاقة بين مرشح الصور ومستخدم الصور الرقمية

يمكن لمحرك بحث الصور تحديد درجة التفاعلية بين مستخدم الويب ومرشح الصور داخل قاعدة البيانات. وكلما كان التفاعل بين مرشح الصور والمستخدم عاليًا، علت درجة تحقيق الصور المسترجعة داخل قاعدة بيانات محرك البحث. ويهدف هذا العنصر إلى قياس المدى الواقعي لهذا التفاعل من خلال التعرف إلى الخيارات المتاحة أمام مستخدمي مرشح الصور ؛ سواء أكانت في تسجيل ملامح البحث أم في ضبط خيارات استدعاء مجموعات الصور. ويسجل الجدول رقم (٥-٢) جوانب التفاعل الممكنة بين مرشح الصور ومستخدمها:-

الجدول رقم (٥-٢)

يوضح العلاقة بين مرشح الصور ومستخدم الصور

محركات البحث		AltaVista	Corbis	Fast	Google
تسجيل ملامح الترشيح	متابعة أثناء البحث	x	✓	x	x
	تلقي الكلمات الدالة	x	x	x	x
	تسجيل ملامح المستفيدين	✓	x	x	✓
التحكم بمرشح البحث	ضبط قوائم التوقف	x	x	x	x
	تحديد مستوى الترشيح	x	x	x	x

يوضح الجدول رقم (٥- ٢) ما يلي :

أولًا ، يعتمد مرشح البحث داخل محرك الصور Fast على أدنى درجات التفاعلية ؛ حيث لا يتيح أي من خيارات التعامل مع المرشح سوى فتح أو غلق عمل مرشح الصور.

ثانيًا : عمل محركي Google, AltaVista على فتح خيارات التفاعل لاختيار شكل المعلومات المراد ترشيحها سواء أكانت المعلومات النصية أم المعلومات المصورة أم كليهما. بينما استخدم Corbis مرشح البحث فقط عند إعادة البحث داخل النتائج.

ثالثًا : افتقرت محركات البحث جميعها إلى مرشح بحث يتيح لمستخدم الصور طرح ملامحه الموضوعية والتفاعل في ضبط سياسات التوقف عن مجموعات الصور. ووضح ذلك في عنصر التحكم بمرشح البحث.

٣/٧/٥ نمط الترشيح المتبع داخل مرشح الصور

تتبع مرشحات الصور طريقتين في ترشيح المعلومات الرقمية المصورة ؛ حيث تستخدم الكلمات الدالة في الترشيح أو أن تعتمد في ترشيحها على تحليل عناصر البنية والألوان. ويهدف هذا العنصر إلي التعرف إلى واقع الترشيح داخل محركات بحث العينة. كما يوضحها الجدول رقم (٥- ٣) :-

الجدول رقم (٥- ٣)

تقنية الترشيح داخل مرشح الصور

محركات البحث		AltaVista	Corbis	Fast	Google
الترشيح النصي	إيقاف الكلمات أو الميتاداتا	✓	✓	✓	✓
	إيقاف الروابط	x	x	x	x
	ملامح المستفيدين	x	x	x	x
الترشيح التحليلي	**تحليل الألوان**	x	x	x	x
	قيم الألوان	x	x	x	x
	الشكل	x	x	x	x
	بنية الصورة	x	x	x	x

كما يوضح الجدول رقم (٥- ٣) فإن محركات البحث تعتمد في ترشيح الصور فقط على الكلمات المفتاحية الواردة ضمن متن أو حقول صفحة الويب. وقد افتقرت مرشحات الصور إلى تبني أي من تقنيات الترشيح بشكلها الكامل؛ حيث تفتقر محركات البحث ذاتها إلى الكفاية في تكشيف ومعالجة الصور الرقمية.

٥/٧/٤ مرشح الصور بين مراحل عمل محرك الصور

تختلف مرحلة عمل مرشح البحث داخل محركات الصور بحسب كيفية توظيف مرشح الصور. حيث يمكن لمحرك البحث وضع سياسة ثابتة في بناء وتنمية مجموعاته من الصور الرقمية، وعلى ذلك فإن أفضل مرحلة لتوظيف مرشح البحث هي أثناء الإضافة وخلال عمل برنامج الزاحف في تصفح صور الويب. وهو الأسلوب الواجب اتباعه في محركات البحث ذات الثقافة العربية والإسلامية، حيث توجد القيم الثابتة غير القابلة للتحديث والتغيير. أما إذا أراد محرك البحث ترك سياسات الترشيح لمستخدمي الصور وتلبية مختلف الحاجات الموضوعية لهم، فإن مرشح الصور ينشط فقط في مرحلة البحث واستدعاء مجموعات الصور.

الجدول رقم (٥-٤) مرحلة الترشيح بين مراحل عمل محرك البحث

محركات البحث	AltaVista	Corbis	Fast	Google
الترشيح أثناء الإضافة	x	x	x	x
الترشيح أثناء التكشيف	x	x	x	x
الترشيح أثناء البحث	✓	✓	✓	✓

يوضح الجدول رقم (٥-٤) أن محركات البحث جميعها يتيح تفعيل مرشحات الصور في ضبط مجموعاته إلى مستخدمي الصور الرقمية، بما يمكن هؤلاء المستفيدين من استدعاء مختلف موضوعات الصور الرقمية. ويمكن لمحرك البحث اتباع أسلوب الترشيح أثناء التكشيف من خلال مضاهاة ملامح المستفيدين الموضوعية والكلمات الدالة على الترشيح بالكلمات الدالة على محتوى الصور الرقمية أثناء تنظيم مجموعات الصور داخل قاعدة البيانات.

٥/٧/٥ فاعلية مرشحات الصور داخل محركات البحث

إن وجود مرشح بحث داخل محركات البحث العربية ليس هو الغاية في حد ذاته، إنما غاية الترشيح هو إثبات فاعلية غلق أو وقف مجموعات الصور غير الملائمة لملامح مستخدمي محرك البحث. وقد عمد الباحث إلى طرح مجموعة من المصطلحات على مرشحات الصور مثل "sex, sexual" ؛ وهي أكثر المصطلحات رفضا داخل البيئة العربية والإسلامية في استرجاع الصور. وذلك لاكتشاف مدى فاعلية تلك المرشحات في استدعاء الصور. واختار الباحث الصفحة الأولى من صفحات عرض النتائج لفحص دقة مرشح البحث في منع محتوى الصور الرقمية:-

الجدول رقم (٥-٥) مستوى فاعلية مرشحات الصور

محركات البحث	AltaVista	Corbis	Fast	Google
ترشيح عالي الفاعلية	✓	x	x	x
ترشيح متوسط الفاعلية	x	✓	x	✓
ترشيح غير فعال	x	x	✓	x

يوضح الجدول جدول رقم (٥- ٥) المستويات المختلفة لفاعلية مرشحات البحث، وذلك كما يلي:-

أولًا : افتقر محرك البحث fast إلى فاعلية مرشح الصور بالدرجة التي تساوت بها طبيعة الصور الرقمية المسترجعة قبل وبعد تنشيط مرشح الصور به.

ثانيًا : حاولت مرشحات البحث داخل Google, Corbis غلق الوصول إلى محتويات الصور الرقمية الصريحة explicit images، غير أنها استدعت محتويات الصور التي تقترب من هذا الموضوع. ويرى الباحث أن مثل هذا الترشيح بلغ من النقص ما يساوي به فشل مرشح البحث في غلق الوصول إلى هذه المصطلحات كاملة، خاصة وأن هذه المحركات تعتمد على الترشيح بالنص باستخدام الكلمات.

ثالثًا : تميز مرشح الصور داخل محرك البحث AltaVista ؛ حيث أغلق مرشح البحث تماما الوصول إلى أي نتائج تتعلق بهذا المحتوى. وكانت نتائجه صفرية، فضلا عن تقرير محرك البحث عن كون هذا الموضوع مغلقًا إلى أن يتم تغيير خصائص البحث مرة أخرى.

٨/٥ العلاقة بين الترشيح ومعاملات نظم الاسترجاع

إن تأثير مرشحات البحث على مجموعات الصور المسترجعة ينعكس بدوره على قيمة معاملات نظم الاسترجاع مثل الاستدعاء والتحقيق والفقد. ويأتي ذلك من تحكم مرشح البحث بكم وكيف مجموعة الصور المسترجعة داخل قاعدة البيانات. ويرى الباحث أن تأثير مرشح الصور على معاملات الاسترجاع يتوقف على مرحلة استخدام مرشح الصور داخل محرك البحث.وذلك كما يلي :-

١/٨/٥ أولا: في مرحلة الإضافة والتنظيم :

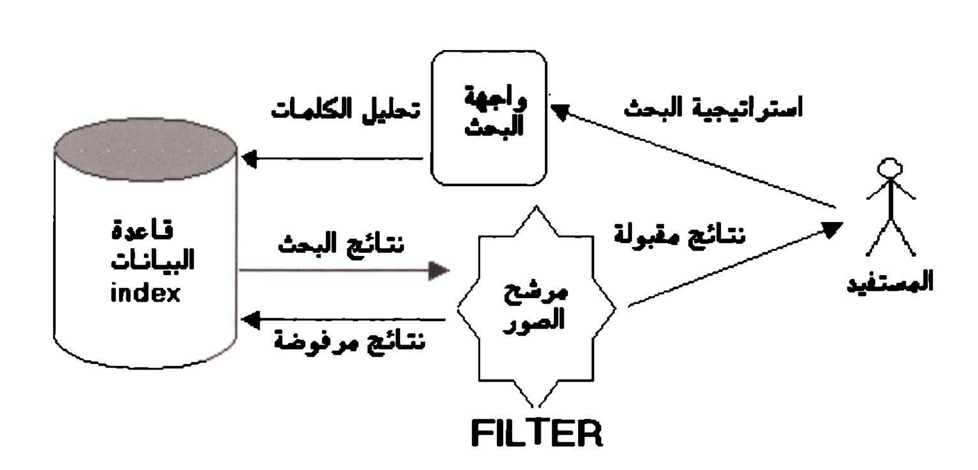
حيث لا يوجد لمرشح الصور أدنى تأثير على معاملات الاسترجاع داخل قاعدة بيانات المحرك ؛ ذلك لأن دور مرشح البحث ينتهي عند إضافة الصورة وتنظيمها داخل قاعدة البيانات، أما معاملات البحث فهي تأتي مع المستفيد في مرحلة بحث مجموعات الصور.

٢/٨/٥ ثانيًا: في مرحلة البحث :

حيث يمكن لمرشح الصور التدخل في كم وطبيعة مجموعات الصور المسترجعة بأحد الأسلوبين التاليين:-

١/٢/٨/٥ ترشيح النتائج

وهو أن يترك مرشح الصور إستراتيجية البحث المصاغة من جانب المستفيد تمر إلى قاعدة البيانات بالشكل الذي ارتضاه المستفيد من كلمات البحث، ثم يبدأ عمل مرشح الصور عند استجابة قاعدة البيانات بإخراج مجموعات الصور الرقمية المطابقة لإستراتيجية البحث ؛ حيث يبدأ مرشح الصور بمضاهاة مجموعات الصور المسترجعة بقوائم التوقف عنده أو بتحليل مجموعات الصور التحليل اللوني ليقرر أي من هذه الصور يمر إلى واجهة عرض النتائج، وأي منها يتم استبعاده. وهذا المستبعد يمكن للباحث أن يطلق عليه نسبة الفقد من مجموعات الصور لكونها طابقت احتياجات المستفيد الموضوعية وخالفت سياسة الترشيح داخل بيئة محرك البحث. ويوضح الشكل رقم (٥-٧) دور وموقع مرشح الصور :-

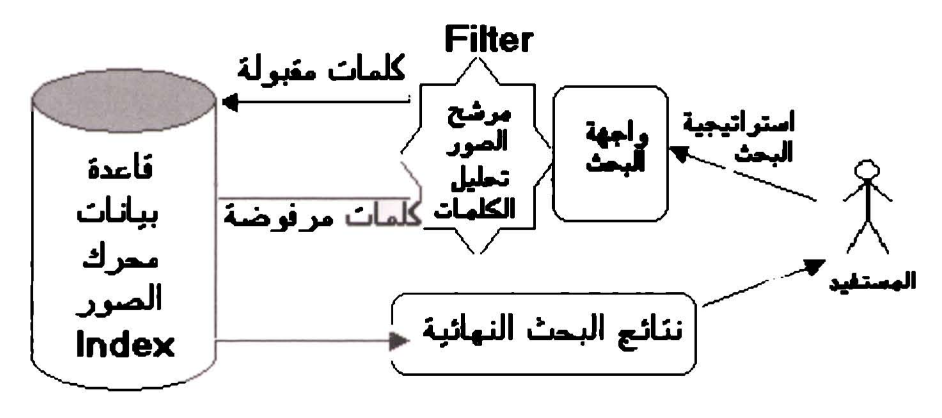


الشكل رقم (٥- ٧)

مرحلة مقترحة لعمل مرشح الصور في ترشيح نتائج البحث

٥/٨/٢/٢ ترشيح إستراتيجية البحث

وهو أن يتحكم مرشح الصور بضبط كلمات البحث وإعادة صياغة إستراتيجية بحث المستفيد قبل تمريرها إلى قاعدة البيانات، وعلى ذلك تصل كلمات البحث إلى قاعدة البيانات بما يوافق سياسة الترشيح داخل محرك بحث الصور، ومن ثم تستدعي قاعدة البيانات مجموعات الصور التي أقرها مرشح الصور. ويوضح الشكل رقم (٥-٨) دور وموقع مرشح الصور:-



الشكل رقم (٥-٨)

مرحلة مقترحة لعمل مرشح الصور في ترشيح كلمات البحث

تؤثر مرحلة الترشيح على مجموعات الصور المسترجعة بفقد نسبة من الصور المستدعاة المتصلة بموضوع المستفيد،غير أن الهدف الحقيقي وراء تفعيل مرشحات الصور هو رسم سياسة لعمل محرك الصور في ضوء ملامح ثقافية يجب اتباعها، دون النظر إلى ما يخلفه أداء مرشح البحث من فقد لمعلومات تتصل بموضوع البحث.

٥/٩ الخاتمة

تهدف محركات بحث الصور الرقمية إلى تلبية احتياجات المستفيدين من المعلومات المصورة، غير أن الصور الرقمية باعتبارها مصدر معلومات يحتاج إلى ضبط وتحكم بما يلائم طبيعة المستفيدين والبيئة التي يعمل بها محرك بحث

الصور. وتستخدم محركات الصور مرشحات الصور كبرامج مساعدة على تنقية وترشيح المعلومات المصورة داخل قاعدة بيانات محرك البحث. وتنتهج مرشحات الصور أساليب مختلفة لترشيح الصور باستخدام الترشيح النصي أو الترشيح التقني بتحليل المحتوى. وتحتاج مرشحات الصور داخل محركات الصور العربية إلى مجموعة محددة من الخصائص لكي يتماشى أداؤها وطبيعة الثقافة الإسلامية. وقد حاول هذا الفصل مناقشة خصائص وعمل مرشح الصور.



الفصل السادس

مواصفات محرك بحث صور يدعم اللغة العربيـة

مواصفات محرك بحث صور
يدعم اللغة العربية

٦/٠ التمهيد

عرضت الدراسة في الفصول السابقة الجوانب والأركان التي يقوم عليها محرك بحث الصور من حيث الإضافة والإعداد الفني ومرحلة البحث والاسترجاع. ويأتي هذا الفصل من الدراسة كثمرة لدراسة بحث واسترجاع الصور الرقمية، هذا إلى جانب التعرف إلى المعالجة الملائمة لخصائص اللغة العربية. ويعتمد الباحث في هذا الفصل في وضع مواصفات محرك بحث الصور الثابتة الداعم للبحث باللغة العربية على النموذج الموضح لمراحل العمل داخل محركات البحث، والذي وضعه Ibrahim A. Al-Kharashi في إطار الحديث عن مشكلات استرجاع اللغة العربية في بيئة الويب[١]. ويقدم الباحث هذا الشكل باعتباره نموذجًا مرئيًا لمحرك بحث الصور تتوفر فيه مجموعة الخصائص المطلوبة لصياغة محرك بحث صور ثابتة داعم للغة العربية. وبالنظر إلى ذلك الشكل يمكن استيضاح الخطوات التالية لمحرك البحث المقترح. ويتكون سياق عرض محرك بحث الصور من العناصر التالية :-

٦/١ مراحل تكشيف الصور الرقمية

يبنى عمل مؤسسات المعلومات على اختلاف درجاتها (المكتبات، المكتبات المتخصصة، مراكز المعلومات)، أو على اختلاف بيئاتها (مراصد البيانات أو

(١) Al-Kharashi, Ibrahim A. A Web Search Engine for Indexing, Searching and Publishing Arabic Bibliographic Databases, King Abdul Aziz City for Science and Technology, Computer and Electronics Research Institute, [2001?] visited 31/12/2002, http://www.isoc.org/inet99/proceedings/posters/085/

محركات بحث الويب) جميعًا على القيام بدور الوسيط بين المرسل والمتلقي في إتمام عملية الاتصال. وتحتاج البيئة الرقمية إلى التنظيم الفني لوسائط المعلومات تماما كما تحتاج إليه المكتبات ومراكز المعلومات، ويمكن القول إن الفارق بين البيئة التقليدية والبيئة الرقمية هو فارق في الدرجة وليس في النوع؛ حيث تهتم محركات البحث بمحتوى صفحات الويب أكثر من اهتمامها بالوسيط كما هو الحال في عمل المكتبات.

١/١/٦ الإعداد الفني للصور الرقمية

إن اختلاف وسيط المعلومات في الصورة عنه في النص جعل من الضروري عند التعامل الببليوجرافي مع الصور الرقمية تحويلها إلى وسيط النص المعتمد على الكلمة. ويرجع ذلك إلى أن تنظيم واسترجاع النصوص داخل قواعد بيانات محركات البحث يعتمد على استرجاع صفحات الويب بالكلمة، غير أن التعامل مع تكشيف الصورة بطبيعتها دون تحويلها إلى وسيط النص الداعم للتحليل الموضوعي عامة والتكشيف خاصة، سوف يحيد بالصور الرقمية إلى تكوين نظم استرجاع تقنية أو فنية لا تتصل من قريب أو بعيد بعمل اختصاصيي المعلومات في تنظيم المعرفة. وعلى هذا كان حرص الباحث على اقتراح مجموعة من حقول البيانات الببليوجرافية الواصفة ماديا للصور الرقمية اعتمادًا على معيار دبلن كور لوصف ملفات الصور الرقمية، وهو ما جاء في الفصل الثاني، ثم العمل على تحديد خطوات لوصف محتوى الصورة باستخدام التكشيف.

ويرتبط بما جاء هنا الحاجة إلى الاهتمام بنتاج ملفات الصور الرقمية ومراعاة إدخال مجموعة حقول الوصف الببليوجرافي الواصفة للصورة والمعبرة عن المحتوى الموضوعي لها؛ حيث إن محرك البحث لا يوفر تلك الخاصية لملفات الصور فهي لا تكشف إلا ما تجده من حقول أو بيانات عن الصورة (الميتاداتا). وقد ذكر Anne Callery وهو أحد المفهرسين داخل دليل البحث Yahoo اعتماد دليل البحث على مجموعة من المعايير لوصف صفحات الويب؛ حيث تخضع صفحات الويب إلى تحليل مادي وموضوعي من جانب هؤلاء المفهرسين لاستخراج أهم الموضوعات التي تعالجها صفحة الويب ومن ثم التعبير عن تلك الموضوعات بالكلمات المفتاحية الملائمة للموضوع، هذا فضلا عن استخراج بعض البيانات

الواصفة مثل العنوان ومسار صفحة الويب والمؤلف... إلخ، وذلك للسماح بالحد الأدنى من وصف صفحات الويب.(١)

وتعرض العناصر التالية المقترحات الخاصة بشأن الإعداد الفني للصور الرقمية على الويب داخل محركات بحث الصور في اللغة العربية:-

٦/١/١/١ تحليل ملف الصورة إلى عناصر محددة

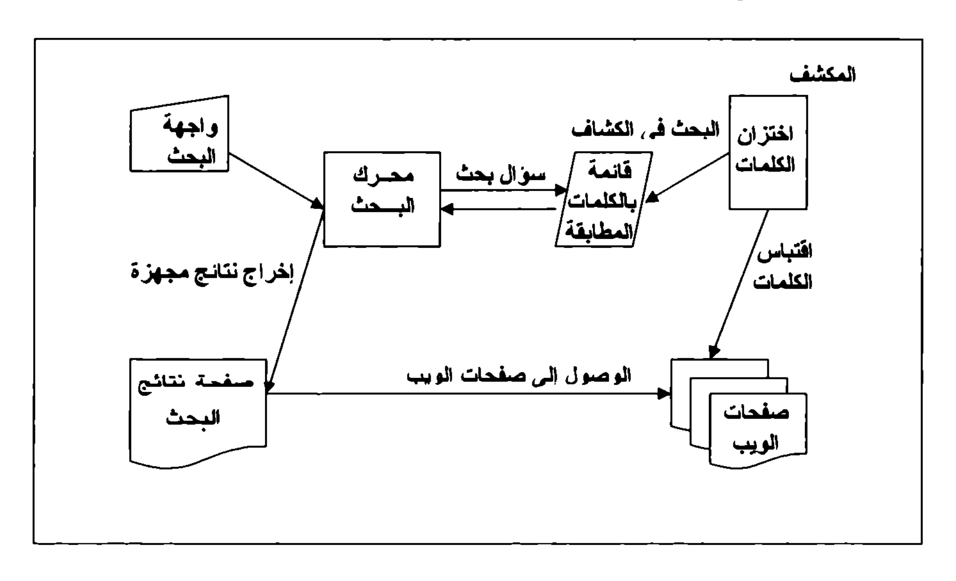
يتركز الحديث في دراسات فهرسة مصادر الويب على العناصر الأساسية للوصف الببليوجرافي لهذه المصادر؛ مثل ما ذكرته قواعد الفهرسة العالمية لملفات الحاسب International Standard Bibliographic Description for Computer files [ISBD (CF)] والتي تم تسميتها في الطبعة الثانية لها باسم International Standard Bibliographic Description for Electronic Resources [ISBD (ER)]، إلا أن مثل هذه القواعد تتعامل مع صفحات الويب بشكلها العام ولا تختص هذه القواعد بوصف ملفات الصور أو الفيديو أو الصوت، ويختلف الوصف الببليوجرافي من وسيط إلى آخر تبعًا لاختلاف الشكل المادي، مما يدل على احتياج ملفات الصور إلى قواعد جديدة للتعامل مع شكل وسيط الصور الرقمية الثابتة.

ويعمل برنامج المكشف على تحليل الصفحة المصدرية للصورة للخروج بالكلمات المفتاحية التي تصف محتوى هذه الصورة؛ إلا أن الوصف الببليوجرافي للصورة الرقمية يجب ألا يتعدى ملف الصورة ذاتها التي تكون الصورة فيها أحد أجزاء هذا الملف. وتقف هذه المرحلة عند تعرف برنامج المكشف على مجموعة من البيانات الببليوجرافية التي يتوجه إليها البرنامج بحسب أماكن سابقة التصميم في ملف الصورة الرقمية. وقد حدد معيار دبلن كور لوصف الصور مجموعة من العناصر المكونة للتسجيلة الببليوجرافية، إلا أنه توجد بعض العناصر الأخرى المهمة في تحديد خصائص الصورة مثل درجة الوضوح ودرجة اللون "كمية وحدات البايت". وتخرج مثل هذه العناصر جميعًا باعتبارها رموزًا واصفة للشكل المادي للصورة الرقمية يتم استخدامه وفقا لما أقرته التعليمات داخل برنامج المكشف. ويجب أن يعمل المكشف على استخدام محتوى الصفحة المصدرية بجانب ملف

---

(١) Callery, Anne. Yahoo! Cataloging the Web, Yahoo! Inc., 1996, visited at 20/10/2003, http://www.library.ucsb.edu/untangle/callery.html

الصورة لإخراج العناصر الواصفة للصورة الرقمية الوصف المادي. ويوضح الشكل رقم (٦-١) لعمل محركات البحث أن صفحة الويب تشكل في مجملها البيانات الخام التي يعمل على تنظيمها وتحليلها برنامج المكشف، وتأتي الصلة بين الوصف المادي ومحتوى الصفحة المصدرية فقط في التعرف إلى المواصفات المادية للصورة مع الحقول الخاصة بالصفحة مثل المسار والعنوان.



الشكل رقم (٦-١)

يوضح مراحل عمل محرك بحث الصور الرقمية

وتعتمد محركات البحث العامة أو المتخصصة على البنية الثابتة لصفحات الويب في ضم وإعداد هذه الصفحات الإعداد الفني، ومن ثم فإن برنامج المكشف يجب أن يملك التعليمات المحددة لأماكن البيانات الببليوجرافية التي يتم تجهيزها في المرحلة التالية لوضعها في شكل الحقل الببليوجرافي. ويتمثل ذلك في حقول اللغة المعيارية HTML meta tags الواصفة لمحتوى وشكل الصورة.

والجدير بالذكر أنه في هذا المرحلة تتحد بنية صفحات الويب العربية والإنجليزية في استخدام حقول ومحارف اللغة الإنجليزية. و يرجع ذلك إلى عدم تعريب لعناصر برمجة اللغة المعيارية HTML, XML؛ حيث لا يوجد في بناء الصفحات العربية ما يقابل عناصر مثل : <Head>, <Title>,<Body> أو غيرها من العناصر

المستخدمة في تصميم صفحات الويب؛ إلا أن التقنيات التي يمكن استخدامها داخل محرك الصور العربي قد تجعل من السهل إدخال النصوص العربية في عناصر مثل Meta Keyword, Meta Title مع الإبقاء على رؤوس هذه العناصر بلغتها الأصلية، و مثالا على ذلك لحقول صورة رقمية:-

<meta title> نهر النيل

<meta keyword> المياه، المراكب، النهر، النيل، مصر، الجمال، الوفاء

٢/١/١/٦ تحويل العناصر إلى حقول ببليوجرافية

تربط هذه المرحلة إخراج العناصر السابقة بمرحلة عمل التسجيلة الببليوجرافية؛ حيث يستكمل برنامج المكشف استخراج كافة العناصر موضع البحث حتى توضع في الشكل المناسب لتفريغها في الاستمارات الآلية ذاتيا داخل قاعدة المكشف. ومن الجدير بالذكر أن الباحث لم يلحظ استخدام كلمة مفهرس بدلا من المكشف في الإنتاج الفكري الأجنبي على الرغم من قيام برنامج المكشف بالتعبير عن الصفحة بإخراج بعض العناصر مثل العنوان ومسار الصفحة؛ حيث بني عمل برنامج المكشف على إخراج الكلمات المفتاحية وتحليل المحتوى دون الوصف الببليوجرافي؛ نظرًا لاهتمام مستخدمي الويب باستدعاء الصفحات بحسب الموضوع دون حقول الوصف المادي.

ويقترح الباحث هنا تغذية محرك بحث الصور بشكل ثابت من ترتيب الحقول الببليوجرافية يتضمن مختلف الحقول الببليوجرافية الواصفة للشكل المادي والمحتوى الموضوعي للصورة. ويقف برنامج المكشف عند تجهيز الحقول الببليوجرافية المعدة للاستقبال في قاعدة التسجيلات الببليوجرافية. ومن ثم يستطيع محرك بحث الصور التعامل مع عناصر ملفات الصور المصاغة باللغة العربية من خلال إدخال نماذج ومحارف اللغة العربية داخل برنامج المكشف في محرك الصور الداعم للغة العربية.

٣/١/١/٦ تكوين التسجيلة الببليوجرافية

بعد أن ينتهي برنامج المكشف إلى إعداد التسجيلة النهائية للصورة الرقمية المكونة من عناصر الوصف المادي والوصف الموضوعي. ويقترح الباحث أن تكون

التسجيلة الببليوجرافية الكاملة للصورة الرقمية العامة في اللغة العربية على الشكل التالي:-

العنوان :

الفنان، المصور (المؤلف) :

الموضوع :

وصف الموضوع (المحتوى) :

تاريخ التقاط الصورة :

تاريخ صنع الصورة الرقمية :

تاريخ وضع التسجيلة :

الناشر :

شكل ملف الصورة :

حجم ملف الصورة :

أبعاد الصورة :

اللون :

تعريف الصورة : مسار الملف :

مصدر الصورة على الويب (في حالة محرك البحث موضع التقاط الصورة من الويب) :

لغة ملف الصورة : في حالة المحركات متعددة اللغات :

تغطية الصورة : المكان :

حقوق نشر الصورة :

ويشير النموذج السابق إلى تعليمات بنية التسجيلة الببليوجرافية، على أن تكون هذه التسجيلات الببليوجرافية في قاعدة بيانات مستقلة تختلف عن تلك التي يتم تجهيز البيانات الخام بها. ويستكمل برنامج المكشف الشق الأول من الإعداد الفني إذا ما كونت التسجيلة الببليوجرافية للصورة الرقمية.

٤/١/١/٦ اختيار الكلمات المفتاحية في ملف الصورة (التكشيف)

لم تكن محركات بحث الويب أو المحددة قد أولت اهتمامًا للوصف البيليوجرافي بقدر ما كان في وصف المحتوى والموضوع. وعلى الرغم من أن الوصف الموضوعي له الأهمية الكبرى في استدعاء صور الويب؛ إلا أن إجراءات الوصف البيليوجرافي تحتاج إلى كفاية أكبر تجاه العناصر المادية وتكوين التسجيلة البيليوجرافية. ويعود هذا إلى استخدام الوصف الموضوعي لطريقة اختيار الكلمات المفتاحية ونقاط الاستدعاء في حالة بقاء النص المصاحب للصورة على ما هو عليه بعد وصف موضوع الصورة.

وتعتمد النظم الآلية في تكشيفها لوسائط المعلومات الرقمية على طريقتين؛ تتمثل الأولى في إخراج الكلمات المفتاحية بأسلوبين إما بالاقتباس Extraction أو بالاشتقاق Derived، أما الأسلوب الأول فيعمل نظام المعلومات في هذه الطريقة على استخراج الكلمات المفتاحية من بين كلمات النص الكامل اعتمادًا على مبدأ التكرار الذي جاءت به الكلمات ذات الصفة المفتاحية على الموضوع؛ ويتم حصر الأشكال المختلفة للكلمة الواحدة مع رد هذه الأشكال إلى الكلمة الأصلية ثم حساب عدد مرات التكرار بين كلمات النص والكلمة الدالة. أما الأسلوب الثاني فهو الاشتقاق ويعتمد على موقع الكلمة في النص كأن تأتي الكلمة المفتاحية في العنوان أو تأتي في العناوين الجانبية أو الكلمات التي تأخذ لونا مختلفًا أو شكلا مختلفًا في الكتابة، مما يفيد معه أن هذه الكلمة ذات تأثير في التعبير عن الموضوع أو محتوى النص المصاحب لوسيط المعلومات، وتتخذ محركات بحث الويب جميعها طريقة الاشتقاق أسلوبا في تكشيف مصادر الويب، النصية وغير النصية؛ إلا أن الأمر يختلف بين نظم معلومات الويب والنظم الأخرى؛ ويعود ذلك إلى اتخاذ محركات البحث من كل كلمة من كلمات النصوص الرقمية مدخلا لاستدعاء مصدر المعلومات مما يعكس اعتماد نظم محركات البحث على الاشتقاق الكامل في تكشيف مصادر معلومات الويب. أما الطريقة الأخرى في التكشيف الآلي فتسمى التكشيف بالتعيين وتتم إجراءات هذه الطريقة من خلال استبدال نظام المعلومات لمجموعة الكلمات الواردة في النص بمجموعة من المصطلحات المختارة التي تعبر عن موضوع مصدر المعلومات الرقمي. ومن شأن هذا النوع من التكشيف الاعتماد على إجراءات أكثر كفاية من تلك التي تعتمد عليها نظم التكشيف بالاشتقاق، نظرًا لأن الواصفات

الموضوعية هنا إنما هي واصفات لموضوع الكلمات الأكثر دلالة على الموضوع داخل النص المكشف.[1]

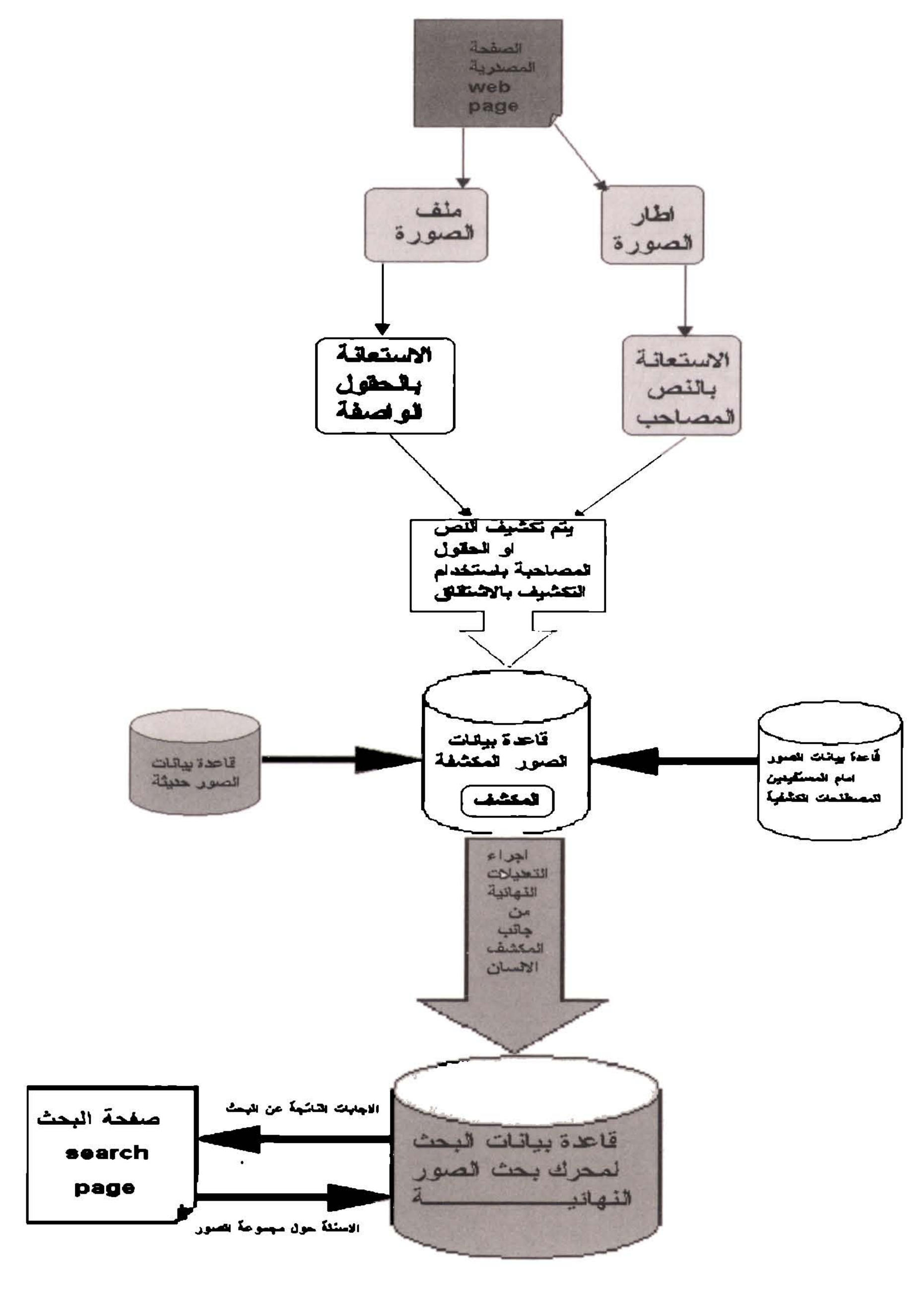
ويجب التأكيد هنا على أن التعامل المعياري وغير المعياري مع ملفات الصور الرقمية يؤثر بدوره على فاعلية التكشيف داخل محرك البحث. حيث يؤدي البناء المعياري لملفات الصور الرقمية إلى وضع ضوابط ومواصفات مقننة من جانب المكشف لاختيار الكلمات المفتاحية داخل ملف الصورة، أما في حالة التعامل غير المعياري وهي حال الويب. الآن فإن برنامج المكشف يلجأ إلى الاعتماد على النص المحيط بإطار الصورة داخل صفحات الويب، مما يترتب على ذلك اختيار أكثر الكلمات دلالة على الموضوع من بين كلمات نص يتسم في ذاته بالضعف في التعبير عن محتوى الصورة. ويقترح الباحث اعتماد الشكل المذكور في الفصل الثاني والمقدم لبناء ملفات الصور الرقمية على الويب؛ حيث يتم تغيير معايير الانتقاء داخل برنامج المكشف اعتمادًا على التقنين الحديث لملفات الصور. ويوضح الشكل رقم (٦-٢) المراحل المقترحة لتكشيف الصور.

ويقترح الباحث اعتبار طريقة الاشتقاق منهجًا لاستخراج الواصفات التي تعبر عن محتوى الصور الرقمية، وذلك اعتمادًا على بنية صفحات لغة الترميز المعيارية HTML، ويكون لموقع كلمات النص الاعتبار الأول في إخراج الكلمات المفتاحية بدلا من مرات التكرار التي جاءت به الكلمات. وتتميز بنية صفحات الويب باحتواء مواقع نصية تمثل نقاط قوة لعرض الكلمات المفتاحية عن المحتوى مثل العنوان أو المسار URL أو حقلي Meta Key word وMeta tags نظرًا لما تحويه هذه الحقول من مصطلحات يعبر بها مصممو صفحات الويب عن الموضوعات التي تتضمنه هذه الصفحات، وقد أكد Robert Corfman على أهمية حقول meta tag في الحصول على الكلمات المفتاحية للتعبير عن محتوى صفحات الويب دون عرضها أمام مستخدمي محرك البحث. وهو ما يعني الاعتماد على بنية مقننة لبناء ملفات صور الويب لكي تتوافق مع إجراءات وتعليمات عمل برنامجي الزاحف والمكشف لالتقاط الصورة الرقمية وتحليلها.[2] ويعد هذا الاعتماد على كلمات الصفحات

---

(١) احمد بدر. **التكشيف والاستخلاص** / تأليف أحمد بدر،محمد فتحي عبد الهادي، ناريمان إسماعيل - - ط.١- . القاهرة، مصر: دار قباء، ٢٠٠١- . ص ص ١٢٦-١٢٨.

(٢) Corfman, Robert. Satisfying the Search Engines - Key Words Meta Tag, the author, 1998, visited6/2/2003,: http://www.corfman.com/robert/CityU/SearchEngine/keywords.htm



المراحل المقترحة لتكشيف الصور داخل محرك بحث الصور

الشكل رقم (٦-٢) يوضح المراحل المقترحة لتكشيف الصور

المصدرية أو الكلمات المصاحبة للصورة أحد ثلاثة أجزاء لتكشيف الصورة؛ حيث يتمثل الجزء الثاني في إشراك المستفيدين لإعطاء الكلمات المفتاحية على محتوى الصورة المعبرة عن آرائهم، أما الجزء الثالث فهو عبارة عن التعديل النهائي من جانب المكشف (الإنسان) لاستكمال المستويات المختلفة لتكشيف الصورة الرقمية. ولقد أكد Imad Hamdan في الحديث عن بناء صفحات HTML أن مسار أو عنوان صفحة الويب يؤدي دورًا مهمًا في تحديد اهتمام محركات البحث بصفحة الويب والتكشيف الخاص بها. وأكد Hamdan على أن مصممي الويب يغفلون أهمية اسم المجال في صفحة الويب؛ حيث يمكن إضافة المصطلحات الواصفة والمحددة لموضوع الصفحة، هذا فضلا عن استخدام اسم المجال الإضافي ومثالا على ذلك فإنه إذا ما أرادت شركة لبيع ذاكرة الحاسب إنشاء صفحة على أن تستخدم مسارًا يحتوي على وصف الشركة مثل www.yoursite.com/computers/memory.html بدلًا من مسار لا يعطي ذلك التعبير مثل www.yoursite.com/Johndoe/mypage.html.[1] ويعكس هذا تأثير عناوين صفحات الويب في الاسترجاع ومن ثم رفع كفاية التحقيق إذا ما اعتمد محرك البحث على مسار الصفحة في التكشيف والاستدعاء والترتيب.

وفي توصية لاحقة لإنشاء صفحات الويب، فقد أكد Hamdan أهمية حقلي العنوان والكلمات المفتاحية في إبراز الكلمات المفتاحية على موضوع \ موضوعات صفحة الويب، كما أن تكرار الكلمات داخل متن صفحة الويب من شأنه التأكيد على فاعلية هذه الكلمات المتكررة في تكشيف صفحة الويب. ويمكن الاستفادة من ذلك في التأكيد على الكلمات المفتاحية التي تعبر عن محتوى الصور الرقمية داخل صفحة الويب؛ حيث تهتم محركات بحث الصور بتكشيف النصوص المحيطة بالصور الفوتوغرافية داخل صفحة الويب.

ويتعلق بالتكشيف الآلي لصور الويب من خلال النصوص مجموعة من العناصر التي يمكن شرحها كما يلي :-

---

(١) Imad Hamdan. Web Site Promotion Techniques for Top Search Engine Listings The Internet Marketing Warriors, 2003, visited at 11/9/2003,

١/٤/١/١/٦ مستويات التكشيف للصور الرقمية

إن العلاقة بين الصورة والنص تسير دائمًا في إطار توضيح الصورة لموضوعات النص وليس العكس، وهذا الأمر هو ما يمكن ملاحظته في الصور الإعلامية المصاحبة للمقالات الصحفية، أو تلك الصور المصاحبة لمتن الكتب أو الصور المصاحبة لمقالات الموسوعات. وتحتاج الصورة عند تكشيفها إلى تبادل الوظائف مع النص لتوفير وصف أكثر دقة لمحتوى الصورة. وتتميز صفحات الويب ذات الألوان المتزاحمة باستخدام الرسومات والصور كدعامة لعرض محتوى النصوص الرقمية؛ أي أن الأمر هنا لا يختلف بين الصور التقليدية والصور الرقمية في بيئة الويب من حيث احتلال الصورة للمرتبة الثانية في عرض محتوى وسيط المعلومات.

وتحتاج الصورة باعتبارها وسيط معلومات مرئيًا إلى ثلاثة مستويات من التحليل الموضوعي؛ وهي أولا مستوى العناصر الأولية (الأساسية) في تكوين الصورة وتمثلها عناصر المحتوى المرئي والتي تقع عليها أعين المستفيدين عند المشاهدة الأولية للصورة، أما المستوى الثاني فهو العناصر الثانوية (خلفية الصورة) وتأتي في العناصر المكملة لشكل الصورة والمساعدة للعناصر الأولية في إعطاء الشكل الكامل للمحتوى، أما المستوى الثالث فهو البعد غير المرئي (المعنوي) للصورة المتمثل في مجموعة معاني تهدف الصورة إلى توصيلها .

وما سبق يفيد احتياج محركات بحث الويب، ذات النظم الآلية في التكشيف، إلى نص أكثر اهتمامًا بمحتوى الصورة وأبعاد المحتوى الثلاثة، مما يحقق للاشتقاق في التكشيف الآلي الكفاية في استخراج الكلمات المفتاحية لمحتوى الصورة متعددة الأبعاد. وقد جاء اقتراح إشراك المستفيدين في إحدى مراحل تكشيف الصور الرقمية مع إدخال التعديلات النهائية من العنصر البشري أمرًا طبيعيًّا في ظل عدم توافر الوصف الكامل لمختلف محتويات وسيط المعلومات المرئي الثابت (الصورة) . ويعتمد اختيار الكلمات المفتاحية في البعد الثالث لمحتوى الصورة على الكلمات الأكثر تكرارًا في التعبير عن معاني صورة محددة، وبذلك يمكن إصباغ تكشيف البعد الثالث لمحتوى الصورة بصبغة التوافق والاطراد بين المستفيدين في عمليتي التنظيم والاستدعاء، حيث يعد البعد الثالث أكثر أوجه

الصورة عرضة لاختلاف آراء المستفيدين مع الأخذ في الاعتبار ثقافة المجتمع وخلفيته في التعبير عن محتوى هذه الصور.

٢/٤/١/١/٦ إجراءات تكشيف ملفات وصفحات صور الويب

يقوم محرك البحث المقترح على أسس تتعدى الأداء العام لمحركات بحث الويب، نظرًا لاعتماد هذا المحرك على مراحل مختلفة لتكشيف الصور الرقمية يمثلها أولا البرنامج الآلي ثم العنصر البشري من جانب المكشف والمستفيدين. وتحاول الكثير من خدمات المعلومات على الويب (محركات البحث) الجمع بين مميزات عمل محرك البحث مع أدلة البحث لتحقيق قدر أكبر من التحقيق للمستفيدين؛ ومثال على ذلك ما عمل عليه دليل البحث Yahoo بالجمع بين البحث الحر للكلمات المفتاحية والإضافة الآلية لقاعدة البيانات مع استخدام الإنسان في تكشيف المواقع لإقامة دليل الموضوعات العامة، وقد ساد لون أدلة البحث الشكل العام للبحث داخل Yahoo، وعلى الجانب الآخر فإن محركات بحث مثل Google عملت على إدخال خدمة دليل البحث ضمن ما تقدمه من خدمات لاستدعاء معلومات الويب مما يعني الكفاية في تنوع خدمات استدعاء المعلومات داخل محركات وأدلة البحث. ويرى الباحث أنه يمكن تشكيل مراحل تكشيف الصور الرقمية داخل محرك البحث المقترح اعتمادًا على إشراك العنصر البشري إلى جانب برنامج المكشف الآلي كما يلي:-

أولًا: يتم تحليل كلمات النص الرقمي في بنية صفحات لغة الترميز المعيارية بواسطة برنامج المكشف لاستخراج الكلمات الأكثر دلالة بحسب أهمية مواقع تلك الكلمات في الصفحات المصدرية للصور. بينما يمكن تعديل تعليمات عمل برنامج المكشف للحصول على جمل بدلا من الكلمات إذا ما تم الاعتماد على النص الشارح لوصف محتوى الصور الرقمية بدلا من الواصفات المفردة والمركبة. ويمكن الاستعانة هنا بما جاء في مقال search engines - submissions advice لتكشيف محركات بحث الويب، حيث أورد المقال احتياج محركات البحث إلى الاهتمام بمجموعة من حقول بناء صفحات الويب مثل <title>.....</title> tag وهو حقل العنوان الذي يحوي مجموعة الكلمات المفتاحية على موضوع صفحات الويب، ويمكن باستخدام ملفات الصور الرقمية المقترحة توجيه برنامج المكشف إلى العمل على

مثل هذه الحقول إلى جانب بعض الحقول المهمة في إخراج الكلمات المفتاحية مثل keywords meta tag الجامع لمصطلحات مصممي صفحات الويب.[١]

والجدير بالذكر أن هذا المقال يشير إلى بعض المشكلات المحتملة في تكشيف صفحات الويب، منها:-

١. تواجه صفحات الويب فترات انتظار طويلة في تكشيف متون تلك الصفحات، وذلك يرجع إلى الأعداد المليونية التي يستقبلها المكشف يوميا.
٢. إن النصوص التي تحويها صفحات الويب في شكل شعارات أو أشكال مصورة لا يتم تكشيفها مع نص اللغة المعيارية.
٣. قد يتطلب تكشيف بعض صفحات الويب التعرف إلى كلمات المرور الخاصة بها.
٤. إن العلامات المختلفة التي توجد في مسارات الويب مثل ؟ \ لا يتم تكشيفها.
٥. قد لا يتم تكشيف الصفحات التي تأخذ وقتا أطول في التحميل، وهو الجانب المحتمل في صفحات عرض الصور الكثيرة.
٦. قد يلجأ بعض مصممي صفحات الويب إلى التحايل على برنامج التكشيف لتسجيل كلمات لا تتعلق بمحتوى صفحة الويب الرئيس.
٧. إن الصفحات التي لا ترتبط مباشرة بالصفحة الرئيسة قد لا يتم تكشيفها.

ويرى الباحث أن مشكلات التكشيف السابقة يجب أن تؤخذ في الاعتبار عند بناء ملفات الصور الرقمية باستخدام أحد المعايير مثل دبلن كور. ومن ثم تفادي عوائق التعامل مع تكشيف بينات تلك الصور الرقمية داخل ملفاتها.

ثانيًا: إتاحة مجموعات الصور الرقمية حديثة الإضافة بالشكل التفاعلي أمام المستفيدين لإبداء المصطلحات أو العبارات المعبرة عن محتوى الصور الرقمية، على أن يكون ذلك في جزء خاص داخل قاعدة البيانات الكبرى لمحرك البحث مع مراعاة إتاحة تلك المجموعات من الصور داخل قاعة البحث الرئيسة لمحرك بحث الصور.

---

(١) search engines - submissions advice. w.p., 2003,visited at 21/6/2003, URL http://alt.venus.co.uk/weed/search/advice.htm

ثالثًا: إدخال التعديلات النهائية من خلال العنصر البشري المتمثلة في تحديد الكلمات المفتاحية الأكثر تكرارا لمجموعات الصور أمام المستفيدين والتعديل في الصياغة النهائية لجمل أو عبارات وصف المحتوى.

ويمكن الاعتماد هنا على ما ذكرته Axandra newsletter حول بعض الملامح الواجب مراعاتها داخل ملفات الصور الرقمية في صفحات الويب لضمان تعامل أكثر جدوى مع برامج الزاحف والمكشف في محركات البحث بصفة عامة؛ إلا أن الباحث قد عمل على تنقيحها بما يتوافق مع بيئة اللغة العربية، وهي كالتالي[1]:-

١. استخدام حقل <IMG SRC = "..."> tag بدلا من تصميم الصورة بملامح لغة "الجافا" JavaScript وذلك لما يتصف به برنامج الزاحف في التعامل مع هذه الحقول وتفقد صفحة الويب من خلالها. على أن تستخدم محارف اللغة العربية لدعم الكلمات المفتاحية في متن ملف الصورة الرقمية.

٢. استخدام الحقول النصية Alt IMG في التعبير عن محتوى الصور الرقمية داخل صفحات الويب من خلال وصف موجز، وذلك إذا احتاجت الصور إلى تقنيات مساعدة للعرض داخل متصفح الويب مثل برامج screen readers. ويفضل هنا استخدام الكلمات المفتاحية والواصفات المحددة للتعبير عن الصور. وبالاعتماد على ذلك، فإنه يمكن صياغة حقول ملفات الصور المنشأة في بيئة اللغة العربية الرقمية بالواصفات المفتاحية على محتوى الصورة الرقمية، حتى وإن فارقت الصورة الرقمية نص صفحة الويب التابعة لها .

٣. استخدام خاصية العنوان <IMG TITLE> داخل ملفات الصورة، وهو ما يعبر عنه بالرسالة التي تظهر أمام المستخدم إذا ما وقف مؤشر الأوامر على الصورة. وهو ما يمثل أحد جوانب صياغة ملفات الصور الرقمية.

٤. تهيئة اسم ملف الصورة Image File Name لدعم برامج الزاحف؛ فبدلا من تسمية ملف الصورة بـ summer.gif فيمكن استخدام water-beach.gif وبذلك يمكن التعامل مع مصطلحات أكثر دلالة. هذا فضلا عن تمييز مسارات الويب بالأدلة مثل water-beach/tourism/people.gif، مما يفيد في التعرف بشكل أفضل

(١) WEEKLY SEARCH ENGINE FACTS. The Axandra newsletter archive, 2003, visited at 1/10/2003,: http://www.axandra.com/news/index.htm

إلى الصورة الرقمية. ويقترح الباحث أن يتضمن حقل العنوان في ملف الصورة الرقمية تعريب مجموعة الكلمات التي وردت في مسار أو عنوان الصورة على الويب. حتى إذا ما أصاب برنامج المكشف هذه الصورة اتضحت الكلمات الأكثر دلالة على المحتوى.

٥. الاهتمام بالنصوص المحيطة بالصورة ؛ حيث تستخدم الكثير من محركات الصور النصوص المحيطة بملف الصورة لاشتقاق الكلمات المفتاحية وهي الأقرب إلى الصحة. وذلك يأتي بتبني معيار ثابت لبناء ملفات الصور الرقمية، على أنه يمكن لملفات الصور الرقمية العربية أخذ المبادرة في بث تلك الملفات المعيارية إلى فضاء الويب مع تخصيص الدعم الآلي اللازم لبحث ووصف اللغة العربية.

٦. استخدام عنوان صفحة الويب؛ بحيث يمكن الاستفادة من عناوين صفحات الويب إذا ما اتصلت الصور الرقمية مباشرة بموضوع الصفحة لاشتقاق الكلمات المفتاحية.

٧. الاهتمام بالروابط بين صفحة الويب والصفحات الأخرى بحيث يتم تحليلها للتعرف إلى الكلمات المهمة في التعبير عن ملفات الصور الرقمية.

٣/٤/١/١/٦ لغات التعبير عن المحتوى والموضوع

يطلق مصطلح اللغة داخل التحليل الموضوعي لوسائط المعلومات على نظم صياغة الواصفات الموضوعية، وتنقسم هذه اللغة إلى لغة طبيعية في حالة عدم استخدام أدوات تقنين أو لغة مقيدة مضبوطة متمثلة في قوائم رؤوس موضوعات عامة أو متخصصة (المكانز) . وتبعًا لاختلاف الشكل بين النص والصورة، فإن الصور الفوتوغرافية تحتاج إلى استخدام اللغة الطبيعية نظرًا لاختلاف وسط حمل المعلومات أيضًا. وتحتاج الصور الفوتوغرافية العامة دون العلمية أو الإعلامية إلى استخدام اللغة الطبيعية في وصف محتوى صور الويب، ويعود ذلك إلى عاملين هما؛ الأول أن محركات بحث الصور تعتمد في تكشيفها على النص الكامل لملف الصورة الرقمية. أو الصفحة المصدرية لاشتقاق الواصفات الموضوعية للصور الرقمية، أما العامل الثاني فهو تميز الصور الفوتوغرافية العامة دون العلمية بوجود البعد الثالث لتحليل محتوى الصورة الذي لا يكاد يوجد في الصور العلمية التي تصور ظواهر طبيعية أو تكنولوجية وهي ما تسمى بالعلوم الصلبة Hard Sciences التي تتميز

بالحصر والتحكم، وهو ما يجعل الصورة العلمية لا تعتمد على أي بعد غير الأبعاد المرئية. في حين أن البعد الثالث قد يتوارى داخل الصور الإعلامية خلف الأحداث السياسية والإعلامية... وغيرها.

وعلى ذلك فإن الصور الرقمية العامة تعتمد على النوع الثاني من لغات التكشيف وهي اللغة الطبيعية، ويأتي مصطلح لغات التعبير كتوضيح لأقسام ثلاثة داخل اللغة الطبيعية الحرة هي التكشيف باستخدام المصطلحات المفردة أو العبارات الواصفة أو الجمل والشارحة لمحتوى الصورة الرقمية، وهو ما يتمثل في النص الكامل لجمل وعبارات صفحات الويب المشتملة على الصور الرقمية موضع البحث داخل محركات بحث الصور، وتعد الجمل الشارحة أكثر لغات التعبير عن محتوى الصور استخدامًا داخل محركات بحث الصور نظرًا لما يقوم به محرك البحث من استدعاء الصور اعتمادًا على كل كلمات نص صفحة الويب. ويرى الباحث أن الصور الرقمية العامة تحتاج إلى أعداد كبيرة من المصطلحات المفردة للتعبير عن عناصر المحتوى الداخلي للصور العامة بأبعاده الثلاثة إذا ما تم الاعتماد على الواصفات الموضوعية. وتمر عملية التكشيف المقترحة لصور الويب داخل محركات بحث الصور بثلاث مراحل تبدأ بعمل برنامج المكشف داخل محرك البحث وتنتهي بالتعديل الإنساني بعد استخدام واصفات مستخدمي محرك البحث المقترحة لتكشيف الصور، ومن شأن تعدد مراحل التكشيف أن يخلف الكثير من المصطلحات المفردة والمركبة المعبرة عن محتوى صور الويب. ويوضح العنصر التالي أهمية الحديث عن اللغة العربية بعد الحديث عن اختيار اللغة الطبيعية لوصف الصور الرقمية.

٥/١/١/٦ الاعتماد على معجم آلي للغة العربية

تتميز اللغة العربية بالكثير من الصفات مثل الاشتقاق والترادف والاشتراك اللفظي... الخ، وكانت تلك الصفات هي ذاتها المشكلات الرئيسة في تحقيق كفاية نظم استرجاع المعلومات باللغة العربية. وقد أورد Badr H. al-Badr المشكلات الرئيسة للغة العربية في جوانب هي[١]:-

---

(١) AL-BADR, Badr H. Using the Internet in Arabic: Problems and Solutions, King Abdul aziz City for Science and Technology, [2002?], visited 22/11/2002,: http://www.isoc.org/inet98/proceedings/5f/5f_1.htm

تبنى اللغة العربية على اشتقاق الأشكال المختلفة للكلمات من الجذور، مما يجعل الاتجاه في معالجة مصادر الويب العربية يعتمد على التحليل الصرفي لجذور الكلمات.

تعتمد معالجة اللغة الطبيعية على قوائم التوقف للكلمات غير الدالة، وتتمثل هذه الكلمات في اللغة الإنجليزية في ألفاظ مثل With, for. أما في اللغة العربية، فإنها تحتوي على اللواصق واللواحق الملتصقة بالكلمات الدالة، علاوة على بعض اللواحق مثل الضمائر.

تتصف اللغة العربية بقدرتها على استيعاب الكلمات الداخلة عليها من اللغات الأخرى، مما يسبب صعوبة في تحليل هذه الكلمات وكتابتها بأشكال مختلفة وهو الأمر الذي يؤثر على كفاية الاستدعاء والتحقيق.

ونتيجة لأعداد مواقع الويب العربية الآخذة في الزيادة، فقد تناولت بعض محركات الويب مثل الإدريسي الذي تنتجه شركة صخر وشركة IBM مشكلات الاسترجاع في اللغة العربية للتعامل مع العدد المتزايد من صفحات الويب العربية. ويتركز الحديث هنا حول الاعتماد المقترح لاستخدام المعاجم الآلية في مرحلة التنظيم (التكشيف) التي تتصل باشتقاق الواصفات الموضوعية لصفحات الويب. وتعد المعاجم الآلية العمود الفقري لتنظيم واسترجاع المعلومات العربية حيث يتوفر بذلك إجراءات التنظيم في تحليل الأشكال المختلفة للكلمة الواحدة، ومن ثم يمكن التدقيق في حساب تكرار الكلمة إذا ما اعتمد محرك البحث على اشتقاق الكلمات المفتاحية بحسب تكرارها في صفحة الويب. أما إجراءات الاستدعاء فهي تتمثل في تحليل كلمات البحث المستخدمة من جانب المستفيد لاستدعاء صفحات الويب، مما يساعد على أداء خدمات الاسترجاع في اللغة العربية مثل استدعاء المقابلات أو استدعاء الوحدات التي تحتوي على معنى كلمة البحث أو جذر كلمة البحث بمختلف السوابق واللواحق، وهو ما يتناوله الحديث في التعرف إلى التحليل الصرفي للغة العربية. ويحتاج محرك بحث الصور باللغة العربية إلى أحد المعاجم الآلية للتحكم في مفردات اللغة، على أن يتسم هذا المعجم بتعدد اللغات المقابلة للغة العربية.

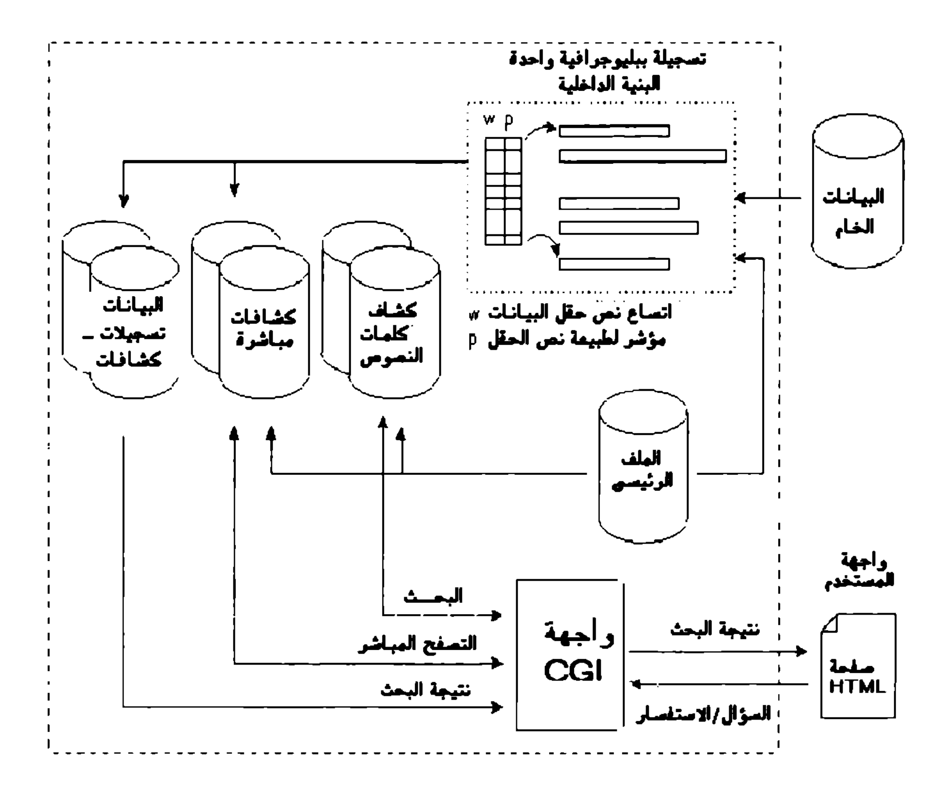
٦/١/١/٦ تكوين قواعد بيانات محرك البحث الداخلية

تأخذ البيانات الخام داخل نظم استرجاع المعلومات مجموعة من الخطوات حتى تصل إلى الشكل المسترجع من نظام المعلومات. وتتمثل البيانات الخام في نظم استرجاع الصور في ملفات الصور وصفحات الويب المصدرية لهذه الصور، وتختلف أوجه التعامل مع هذه الصفحات بحسب مرحلة الإعداد الفني ومما يحتاج معه من توفير قواعد بيانات مختلفة لتخزين الشكل الواحد لكل صفحات الصور داخل نظام المعلومات. ويحتاج برنامج المكشف إلى قاعدة بيانات لتخزين مجموعات النسخ المضافة بواسطة الزاحف لتحليل وتكوين ملفات قاعدة فهرس محرك البحث، وتدخل الصور في مرحلة أخرى هي التفاعل مع مستخدمي محرك بحث الصور لتلقي مجموعة الواصفات الموضوعية المعبرة عن آراء مختلفة تجاه البعد الثالث للصور الرقمية.

ويوضح نظام Al-Kharashi في الشكل رقم (٦-٣) أنواع ملفات قاعدة بيانات محرك البحث المتنوعة بين ملفات البحث داخل التقسيم الموضوعي (الدليل)، أو تلك التي تحتوي على المجموعات المكشفة من صفحات الويب إلى جانب تخزين نسخ هذه الصفحات. وعلى ذلك فإن محرك بحث الصور الرقمية يحتاج إلى أنواع مختلفة من الملفات التي تتيح معالجة الصور في مراحل مختلفة أو ملفات البحث أمام المستفيدين، وقد تتكامل مجموعة من قواعد البيانات تكون المجموعات الكاملة لمحرك البحث. وتأتي الحاجة إلى مجموعة الملفات السابقة بأنواعها نتيجة لطول فترة المعالجة الفنية لصفحات الويب داخل محركات البحث حتى يمكن للمستفيدين الحصول على صفحة الويب كإحدى وحدات نتائج البحث داخل المحرك، وتختلف هذه الفترة بين محرك بحث وآخر، فقد تصل هذه الفترة إلى ثلاثة اشهر داخل محرك البحث Yahoo نظرًا لاستقباله مليون صفحة ويب يوميا يتم إضافتها إلى قاعدة البيانات[١] .

---

(١) Frequently Asked Questions. Web Drive Canada Inc., 2001, visited 30/1/2003, http://www.webdrive.ca/faq.html



الشكل رقم (٦- ٣) يوضح مراحل عمل محرك البحث المقترح

٢/١/٦ إعداد واجهة صفحة التسجيلة النهائية

١/٢/١/٦ شكل التسجيلة النهائية

تنتهي عملية الإعداد الفني بشقيها الوصف المادي والوصف الموضوعي بتكوين شكل تسجيلة الوصف الببليوجرافي الكاملة، وتضم هذه التسجيلة ضمن حقولها الواصفات الموضوعية بجانب الوصف الموضوعي الشارح لموضوع الصورة إذا ما تم استخدام العبارات أو الجمل لتحليل المحتوى. ويختص هذا الجزء بالشكل العام لصفحة التسجيلة النهائية وما يمكن أن تحويه هذه الصفحة بجانب عناصر الوصف الببليوجرافي من واصفات موضوعية إلى عرض الصورة الرقمية بالحجم الملائم داخل التسجيلة. وتنقسم العناصر المكونة للصفحة النهائية إلى حقول واصفة لمحتوى وشكل الصورة وإطار عرض الصورة الرقمية بجانب الخدمات المقدمة لمستخدمي محرك البحث للتعامل مع الصورة المحددة من الوحدات المستدعاة.

ويقترح الباحث تقديم التسجيلة النهائية كاملة الحقول بجانب إطار الصورة الرقمية في القسم الأعلى من صفحة التسجيلة النهائية، على أن توضع الخدمات الممكنة للتعامل مع الصورة في شكل رموز تعبر عن هذه الخدمات مثل شكل ▼ الدال على تنزيل الصورة Loading، إذا لم تخضع هذه الصور بطبيعتها لحقوق النشر. أما القسم الثاني من صفحة التسجيلة النهائية فيمكن أن يحتوي على النسخة الكاملة للصفحة المصدرية للصورة الرقمية موضع الاختيار، حيث يحتاج بعض المستفيدين إلى التعرف إلى محتوى الصورة في الإطار العام لمحتوى صفحة الويب مما يساعد على فهم الموضوع أكثر. ويفصل بين قسمي الصفحة إطار يسمح بالتحرك خلال أي من القسمين العلوي أو استعراض صفحة الويب فقط.

وتتعامل نظم الاسترجاع مع التسجيلات الببليوجرافية بثلاثة مستويات للعرض أمام المستفيدين، وهي تختلف بين المستوى المختصر والكامل والفني؛ حيث يتضمن المستوى المختصر الحقول المهمة في التعريف بوسيط المعلومات، ويتضمن المستوى الكامل الحقول الكاملة، ويزيد على حقول الوصف بيانات أخرى مثل اسم المدخل أو تاريخ الإدخال ويشتمل عليها المستوى الفني في العرض. وهذا يعني أن خيار مستوى العرض يمكن التحكم فيه مع وقف التعديلات لمختلف هذه المستويات. ويأخذ إطار الصورة في حجم متوسط موقعًا موازيًا لحقول الوصف على أن تكون هذه الصورة رابطًا للحجم الكامل للصورة بدون بيانات مصاحبة ويمكن استخدام حقول الوصف الأخرى كروابط استرجاعية للبيانات المعبرة عنها مثل اسم المصور أو تاريخ الصور أو الكلمات المفتاحية داخل حقول الوصف الموضوعي.

٦/١/٢/٢ أنواع الخدمات المقدمة تجاه التعامل مع الصورة

تختلف أنواع الخدمات المقدمة بين محركات البحث تجاه الصور الرقمية حسب طبيعة محركات البحث التجارية المتخصصة أو محركات البحث العامة. فتقدم محركات البحث مثل Corbis خدمات مثل طلب شراء الصورة أو التعرف إلى حقوق النشر أو إرسالها كبطاقة مراسلة أو تحميل الصورة على منفذ المستفيد .

٦/٢ مراحل عمل محرك الصور

يظهر الشكل رقم (٦ - ٣) الذي قدمه Al-Kharashi الإجراءات الآلية لعمل محركات البحث العامة سواء تلك العامة في أشكال المعلومات أو العامة في

موضوعات مقتنياتها؛ إلا أن مخطط الإجراءات السابق قد أظهر ضمنًا عمل برنامج الزاحف الخاص بإجراءات الإضافة وهو الرابط بين البيانات الخام الداخلة إلى قاعدة بيانات المحرك وإفراغ بيانات الموقع المضافة داخل الشكل الببليوجرافي المعد من خلال برنامج المكشف. وهذا يأتي في إطار تركيز الشكل السابق على تحليل البنية الداخلية لمحرك البحث دون التأكيد على الإجراءات التي تحملها الأسهم في انتقالها من مكون إلى آخر داخل بنية محرك البحث. وعمل مخطط Al-Kharashi على إظهار تقسيم قاعدة محرك البحث إلى قواعد ثلاث يختلف شكل البحث داخلها باختلاف شكل البيانات الببليوجرافية وطبيعة الحقول النشطة في البحث أمام المستفيد.

وقد استخلص الباحث المراحل الإجرائية التالية لعمل محرك البحث لإصباغها بصبغة اقتناء ومعالجة وبحث وسيط الصور الفوتوغرافية الرقمية، وهو ما يلي بدوره التركيز على معالجة وسيط معلومات الصور الرقمية على الويب، وكانت المراحل كالتالي:-

١/٢/٦ مرحلة عمل برنامج الزاحف أو العنكبوت

يؤدي برنامج الزاحف في نظام استرجاع الصور دور متخصصي التزويد في أعمال الإضافة والانتقاء، ويتحرك برنامج الزاحف أو العنكبوت متبعًا لأوامر وسياسة محددة في التعامل مع مصادر الصفحات. ويؤدي برنامج العنكبوت كل مهام التزويد من حيث إضافة المصادر والمجموعات الحديثة ومتابعة حالة المجموعات الحالية من إحلال وتجديد لكل التالف والمفقود؛ ويتمثل هذا العمل داخل محركات البحث في الزيارات الدورية التي يقوم بها برنامج الزاحف داخل كل روابط قاعدة البيانات للتعرف إلى الروابط المكسورة broken links أو التحديثات التي مرت بها صفحات الويب سابقة الانضمام لمحرك البحث. ويعتمد برنامج الزاحف في الوصول إلى مواقع وصفحات الويب وضمها إلى قاعدة البيانات يعتمد على طبيعة شبكة الويب، حيث يمكن نظريًا زيارة كل صفحات الويب متخذًا من روابط الصفحات وتداخل هذه الروابط مسارًا واحدًا يمر بكل المواقع والصفحات المنتمية للويب، ويتضح ذلك أكثر في تحليل برنامج الزاحف لكل الروابط التي يجدها في صفحة واحدة تم ضمها إلى قاعدة البيانات، ولعل هذه المهمة تسير ذاتيا اعتمادًا على تعليمات سابقة لبرنامج الزاحف. ويفسر عمل برنامج العنكبوت الأعداد الكبيرة

التي يتم إضافتها يوميًا إلى قاعدة بيانات محرك البحث، والذي يختلف هنا عن أدلة البحث القائمة على استخدام العنصر البشري في إضافة وتحديث المجموعات في قواعد بياناتها .[١]

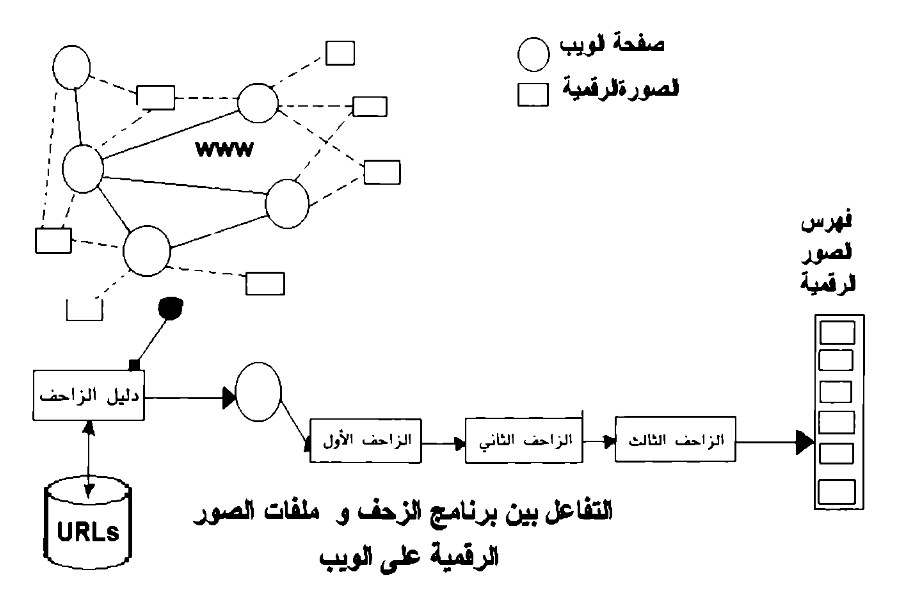
ويوضح العرض السابق مدى التأثير الذي يحدثه برنامج الزاحف في سمات مجموعات محرك الصور إذا ما تم تزويد هذا البرنامج بخصائص محددة مثل قوائم التوقف عن الصور التي لا تناسب طبيعة محددة للاستخدام. ويلتزم برنامج العنكبوت بأداء إجراءات الركن الأول من أركان ثلاثة لعمل محرك بحث الصور، يليها الإعداد الفني للصور ثم الاستدعاء بحسب الحاجة الموضوعية. ويمكن تقسيم المواصفات المقترحة لمهام برنامج الزاحف "اختصاصي التزويد" في ثلاثة عناصر كما يلي :-

١/١/٢/٦ خصائص البحث والتحديث

يقترح الباحث تزويد برنامج الزاحف بمجموعة من التعليمات تمثل سياسة الإضافة والاقتناء، ومن التعليمات؛ أن يجيد برنامج الزاحف التعرف إلى ملفات الصور الرقمية الفوتوغرافية فقط، خلاف ما يكون عليه مصطلح Image من اتساع يشمل اللوحات الفنية والأشكال البيانية... وغيرها من وسائط المعلومات المرئية الثابتة، وهذا ما يجعل مجموعات قاعدة البيانات أكثر إفادة في مرحلة الاستدعاء واسترجاع الصور؛ فمن الطبيعي أن البحث يمكن أن يستدعي معه كل مصادر المعلومات عن موضوع محدد؛ إلا أن طبيعة هذه المصادر وشكل المعالجة لا يعطي الدرجة المطلوبة من التحقيق لهذا الموضوع. وقد يمتد الأمر هنا إلى الاعتماد على تحليل المحتوى Content Retrieval لمساعدة برنامج الزاحف في تحقيق فائدة التزويد. وتشمل سمات التزويد أيضًا التعامل مع ملفات الصورة أو الصفحات المصدرية التي تحوي هذه الصور، ويرجع التعامل مع الصفحات المصدرية فقط إلى التحليل الموضوعي للصور الفوتوغرافية الرقمية الذي يعتمد في بيئة الويب على اشتقاق نقاط الاسترجاع من داخل متن صفحات الويب الحاوية لملفات الصور الرقمية. ويأتي المحور الثاني من عمل الزاحف متمثلا في سمات التحديث وإعادة

(١) How Search Engines Work. Search Engine Watch, 2002, visited at 18/4/2003, http://www.searchenginewatch.com/webmasters/article.php/2168031

الزيارة للصفحات السابق إضافتها، ويأتي ذلك ضمن الهدف من متابعة التحديثات والإضافات لهذه الصفحات، وتختلف طبيعة الصور عن المواد النصية في التحديث؛ حيث يتمثل تحديث النصوص في إضافة أو حذف فقرات أو أجزاء من النص الأصلي، وهو ما لا يحدث مع الصور التي يكون التحديث فيها على مستوى الوسيط بإضافة أو حذف صور بأكملها، ولا تكون بتغيير شكل الصورة في العناصر المرئية. والتحديث لقاعدة الصور هو المعنى الآخر لكلمة الإضافة؛ حيث يكون التحديث لكل مجموعات الصور الرقمية.



الشكل رقم (٦- ٤)

يوضح مراحل التعامل بين الزاحف وصفحات ملفات الصور(١)

يوضح الشكل رقم (٦-٤) أن برنامج الزاحف ينطلق في عمله مع الإضافة من خلال الروابط القائمة بين صفحات الويب وبعضها البعض، وهو الأمر الذي يجعل من اليسير على برنامج الزاحف الإبحار في الأمواج المتزايدة من صفحات

(١) Smith, John R. & Chang, Shih-Fu. Searching for Images and Videos on the World-Wide Web, Department of Electrical Engineering and Center for Image Technology for New Media, Columbia University, New York, 1996, visited at 12/12/2002, http://www.dlib.org/dlib/february97/columbia/02chang.html

الويب. ومع الأخذ في الاعتبار عدم وجود الروابط بين ملفات الصور الرقمية أو روابط الصور، فإن برنامج الزاحف على ذلك لا يستطيع الوصول إلى ملف صور أو صورة رقمية من صورة أخرى، وإنما فقط من خلال صفحة أخرى. وذلك ما أدى إلى استخدام مراحل ثلاث لوصول برنامج الزاحف إلى ملفات الصور ذاتها. وعلى الرغم من أن صفحات الويب هنا ليست الهدف من الإضافة؛ إلا أنه لا يمكن دعم برنامج الزاحف بخاصية إضافة ملفات الصور مباشرة، وهذا يرجع إلى الدور الأساس الذي يقوم به متن صفحة الويب في التكشيف بالاشتقاق للصور الرقمية التي تحويها تلك الصفحات. ويرى الباحث اعتمادًا على ما سبق من أهمية صفحات الويب، أنه يمكن التعامل مع صفحات الويب في مرحلة الإضافة الأولى ثم التعرف إلى روابط الصور الرقمية في تلك الصفحات في المرحلة الثانية، وتكون المرحلة الثالثة بإخراج الصور الرقمية إلى قاعدة بيانات الصور. ويمكن الاستعانة بالجدول التالي الموضح لنهايات مسارات محددة يعمل عليها فقط برنامج الزاحف، مما يمكن معه التأكيد على نوع بعينه من ملفات الويب. وذلك كما يلي :-

الجدول رقم (٦- ١)

يوضح نهايات العمل لبرنامج الزاحف

Extension	Type
.gif	Compuserve image format
.jpg,.jpeg,.jpe,.jfif,.pjpeg,.pjp	JPEG image format
.qt,.mov,.moov	Quicktime video format
.mpeg,.mpg,.mpr,.mpv,.vbs,.mpegv	MPEG video format
.avi	Microsoft video format
.htm,.html	Hypertext Markup Language

٢/١/٢/٦ قائمة توقف عن الصور

يقترح الباحث أن يعتمد برنامج الزاحف على قائمة توقف يمتنع بها عن تلك الصور غير الملائمة لطبيعة مستخدمي الصور. ويأتي ذلك اعتمادًا على مجموعة من التعليمات التي يسير عليها الزاحف في التقاط الصور، وهو ما يتضمن سياسة الترشيح لوسائط المعلومات ليس فقط في الاستدعاء، وإنما عند الإضافة

أيضا، هذا ليحقق محرك الصور القدر الأكبر من التحكم في أوعيته من الصور. وتعطى تعليمات التوقف لبرنامج الزاحف عن صور محددة عندما يحتوي مسار أو عنوان هذه الصور على بعض الكلمات المطابقة لكلمات داخل قائمة التوقف المصاحبة لبرنامج الإضافة.

٣/١/٢/٦ نسخ صفحات الويب

لا يتوقف عمل الزاحف عند الوصول إلى صفحات وصور الويب، وإنما يتعدى ذلك إلى تغذية قاعدة بيانات المكشف بكامل شكل صفحة الويب، مما يعني نسخ هذه الصفحة. يهدف الزاحف من خلال نسخ صفحات الويب إلى تحليل مختلف الروابط links والمسارات URLs داخلها، ومن ثم يستطيع الزاحف الوصول إلى صفحات جديدة وموضوعات أخرى. أما الهدف الآخر من عملية النسخ فيتمثل في توفير كامل النص لبرنامج المكشف وإخراج الأشكال المختلفة من وسائط حمل المعلومات " النص، الصوت، الصورة ". هذا إذا ما كان محرك البحث يدعم البحث في أشكال مختلفة من الوسائط .

تختلف محركات البحث من حيث الاحتفاظ بنسخ صفحات الويب؛ حيث تعتمد بعض المحركات على وضع صفحة الويب كأحد العناصر المكونة للتسجيلة الببليوجرافية، في حين أنه يتم حذفها في بعض المحركات التي تستخدم النسخ فقط لعملية الإعداد الفني ثم يتم حذف النسخ مرة أخرى. وتواجه محركات الصور الرقمية عائق حق النشر في التعامل مع نسخ الصور وهو ما تتصف به الملايين من الصور الرقمية على الويب؛ إلا أن الوصول إلى الشكل الكامل للصور الرقمية يمكن أن يكون في صفحتها المصدرية. وتعوض اللقطات Thumbnails ذلك في بناء التسجيلات الببليوجرافية.

٤/١/٢/٦ معوقات وحلول عمل برنامج الزاحف

إن اعتماد محركات البحث على سياسات مختلفة في التعامل مع صفحات الويب إضافة ومعالجة وبحثًا، جعل الكثير من المعوقات تكتنف عمل برامج الزاحف في التعامل مع صفحات الويب. وقد آثر الباحث إدراج هذه المعوقات في هذا الجزء من الدراسة للاعتناء بالحلول المقترحة في بناء برامج الزاحف، وقد

أشار Ray Denenberg إلى هذه المعوقات موضحًا بعض المقترحات التي يمكن التعامل بها. وهي كما يلي [١]:-

- إن الزاحف لا يعتمد على معيار ثابت في عمق التكشيف لصفحات الويب، وذلك يمكن أن يرجع إلى قيمة صفحة الويب أو كثافة عدد الصفحات داخل موقع الويب أو حقوق الملكية التي تمنع الزاحف من التعامل مع متن صفحة الويب. هذا فضلا عن ارتكاب برامج الزاحف لخطأ إضافة الصفحات الاحتياطية mirror sites لمواقع الويب، وتتسم هذه الصفحات بالبعد عن التحديث مما يترتب عليه تكدس قاعدة بيانات محرك البحث بإصابات خاطئة في عمليات البحث.
- إن اعتماد برامج الزاحف على استخدام الروابط للوصول إلى الصفحات يجعل هذه البرامج تعمد إلى إدخال أنواع من المعلومات التي لا يمكن تكشيفها داخل محرك البحث مثل ملفات الصوت والصورة التي يصعب على بعض محركات البحث التعامل معها.
- تعمل برامج الزاحف دون النظر إلى أهمية صفحات الويب التي يتم إضافتها إلى قاعدة بيانات المحرك، ذلك لما تفقده هذه البرامج من المهارة في تقييم المعلومات داخل صفحات الويب.
- تواجه برامج الزاحف مشكلة أخرى تتعلق باستتار المعلومات داخل الويب بعيدًا عن أيدي المستفيدين،و ذلك يرجع إلى عدم إتاحة المعلومات داخل الخادمات التي تحملها نظرًا لمنع الإتاحة لمصادرها، نزولًا عن المدة الزمنية التي تمر بها صفحات الويب منذ بداية تحميلها على الويب إلى تعرف الزاحف عليها وتكشيفها داخل محرك البحث.

الحلول المطروحة :-

- إن الوقوف على الحلول الكافية للمشكلات السابقة يأتي من خلال التعاون بين القائمين على إعداد الكشافات ومنشئي مصادر المعلومات من المؤلفين والناشرين، ويأتي الحل الأول في شكل تجزئة العمل بين برامج الزاحف

---

(١) Denenberg, Ray. Structuring and Indexing the Internet, Library of Congress, 1996, visited 1/11/2003, http://lcweb.loc.gov/z3950/agency/papers/italy.html

المختلفة، مما يؤدي بدوره إلى تحسين العمل في إضافة وتحليل مواقع الويب، على حين أن هذا الاقتراح يصعب تنفيذه من الناحية العملية لما يؤثر على الجوانب المادية وإدارة الأعمال لمحركات البحث.

- يتمثل الحل الثاني في تمكين محرك البحث ذاته من التعامل مع صفحات الويب القادمة من الزاحف بتحديد قيمة هذه الصفحات وتحديد اللائق للتكشيف منها والامتناع عن الأخرى، وهذا يمكن محرك البحث من التحكم في قيمة المعلومات المكشفة والتعرف إلى إحصائيات إدخال وبحث المعلومات داخل قاعدة البيانات.
- ويمكن طرح حل آخر وهو قيام منتجي مصادر المعلومات من مؤلفين وناشرين بإعداد حقول البيانات الخلفية التي ترفق مع صفحات الويب بما يمكن للمكشف التعامل مع محتوى صفحة الويب وتعرف برنامج الزاحف على ماهيتها؛ إلا أن ذلك يكتنفه بعض المعوقات مثل عدم توافر القدرات الكافية لدى الكثير من منتجي المعلومات على الويب لإعداد حقول البيانات الخلفيةث إضافة إلى تعارض الشكل الموحد لكتاب حقول البيانات مع السياسات المختلفة للتكشيف داخل محركات البحث؛ حيث توجد الكثير من الآليات التي تعمل بها محركات البحث مثل التكشيف بتاريخ إعداد صفحة الويب، تاريخ التحديث الأخير لها، وهي خلاف الآليات الأخرى غير المعروفة أو التي يتم البحث لتحديثها الآن.

وأخيرًا يمكن القول بأن التعامل مع مشكلات برنامج الزاحف يمكن التعامل معها جذريا من خلال التوجه إلى معايير موحدة تخدم عمل محركات البحث المختلفة على الويب، وهو ما تحتاج محركات البحث للعمل به خلاف تبني سياسات مختلفة يتميز بها محرك بحث عن آخر.

٦/٢/٢ مرحلة البحث في المحركات

يتكامل نظام بحث الصور الرقمية مع نظامي الإضافة (التزويد) والإعداد الفني (المعالجة)؛ لتحقيق نظام متكامل لاسترجاع الصور الرقمية. ويستند نظام بحث الصور داخل محركات البحث العربية اعتمادًا على عناصر فرعية مثل؛ آليات البحث الملائمة للغة العربية، والمعاملات المكونة لآليات البحث، وأدوات معالجة

استرجاع اللغة الطبيعية للنصوص العربية، بجانب الشكل العام لبناء نظام الاسترجاع سواء التصفح بالتقسيم الموضوعي أو استخدام البحث الحر بالكلمات المفتاحية. وتتناول الدراسة في العناصر التالية نظام بحث الصور الرقمية ذات النصوص العربية:-

١/٢/٢/٦ استخدام أسلوب التصفح

تحتاج الصور الرقمية في استرجاع مجموعاتها إلى تقسيمات موضوعية أكثر دقة من تلك التي تخدم استرجاع النصوص؛ حيث تتميز الصور بالبعد الثالث للمحتوى (المعاني) مما يحتاج إلى دقة أكبر في ربط الصورة الرقمية بالمصطلح المحدد في قوائم المصطلحات الموضوعية. ويزداد الأمر أهمية عندما تكون لغة الاستدعاء هي اللغة العربية بوفرة مترادفاتها وتراكيبها المتداخلة مع المشتركات اللفظية... إلخ، وتحتاج الصور الرقمية إلى أداة مقننة يمكن أن تتمثل في قوائم رؤوس الموضوعات أو الكشاف النسبي التابع لخطة ديوي الحصرية؛ إلا أن الاعتماد هنا يكون على المصطلحات وليس الأرقام المقابلة لها. وتحتاج قواعد الصور الرقمية المتخصصة بطبيعتها إلى أدوات المكانز خلاف ما تتسم به قوائم رؤوس الموضوعات من حصر الموضوعات والمصطلحات المتنوعة[١].

ويستخدم أسلوب التصفح أكثر من شكل لعرض محتويات قاعدة بيانات محرك الصور من أشهرها الاعتماد على المصطلحات فقط للتعبير عن الموضوعات العامة والمتخصصة، أو أن يتم استخدام الصور كممثل لموضوعاتها ترمز كل وحدة منها إلى موضوع محدد عام أو خاص. ويقترح الباحث استخدام أداة حصر لرؤوس الموضوعات العامة في شكل الألفاظ على أن تتداعى هذه المصطلحات في قوائم وتقسيمات من العام إلى الخاص إلى الأكثر تخصصا، يمكن معها ترميز المجموعات الكاملة لقاعدة بيانات الصور الرقمية. وحري بالذكر أن أسلوب التصفح يعمل بقدر أعلى من الكفاية كلما تحددت مجموعات الصور، وهو ما لا تتسم به قواعد محركات الصور على الويب؛ نظرًا لاعتماد محركات البحث على برنامج الزاحف الآلي مما يمكن إضافة الملايين من ملفات الصور، وينتهي هذا

---

(١) Subject Tree Searches: Intro To Subject Trees.[with out publisher], 2003?, visited 28/8/2003, http://www.concentric.net/~Rkriesel/Search/S-Trees.html

الأمر إلى تراكم مجموعات الصور الرقمية داخل محركات البحث يصل إلى مئات الملايين من ملفات الصور، وقد استخدمت مقالة C. Olivia Frost <http:// www.si.umich.edu/~cfrost/index.html>) السابقة أداة ضبط المصطلحات (Art and Architecture Thesaurus AAT لتكشيف الصور الرقمية في قاعدة الفنون والعمارة، إلى جانب نظم تكشيف أخرى مثل Library of Congress Thesaurus for Graphic Materials LCTGM أيضًا الأداة المتخصصة Thesaurus of Geographic Names TGN لتمثيل صور الأشخاص في الجغرافيا.[١]

ويقترح الباحث الاعتماد على بناء هذه القوائم من طبيعة مجموعات الصور داخل قاعدة بيانات محرك الصور وذلك لما يمكن معه توفير الجهد المستنفذ في تهذيب مصطلحات تلك القائمة المعدة مسبقًا، ويمكن بناء تلك القائمة من خلال أسلوب التعامل الآلي مع بناء المكانز Thesaurus؛ حيث تعمل محركات البحث أثناء عملية التكشيف على اختيار أعلى المصطلحات دلالة داخل صفحة الويب ثم تضمينها في الترتيب الملائم داخل قائمة التصفح بإشراف العامل الإنساني داخل محرك البحث، على أن تكون تلك المصطلحات محدودة العدد بحيث لا يتعدى عددها للصورة الواحدة أكثر من مصطلحين أو ثلاثة. كما أن تحديث تلك القائمة يتم بشكل آلي ومساير لتطور المصطلحات المستخدمة مع الصور الفوتوغرافية العامة على الويب.

يمكن أيضًا من خلال التعرف إلى مميزات وعيوب أسلوب التصفح الوصول إلى الشكل المرضي للمستفيدين لاستخدام هذا الأسلوب في محرك بحث الصور المقترح، ويمكن تناولها على النحو التالي :-

مميزات أسلوب التصفح

١) يتميز أسلوب التصفح عن أسلوب الكلمات المفتاحية بتحقيق التفاعلية بين النظام والمستفيد؛ حيث يجد المستفيد في بحثه خلال قوائم المصطلحات إشارات ترشده إلى الوقوف على ما يريده من معلومات و مساعدته في الوصول إليها.

---

(١) Frost, C. Olivia and Noakes, Anna.optc.

٢) يعمل أسلوب التصفح أكثر على تصحيح مسار المستفيد في الوصول إلى المعلومات، ذلك من خلال تجهيز و إعداد المصطلحات أمام المستفيد دون الحاجة إلى محاولات الصواب و الخطأ في صياغة كلمات البحث.

٣) لا يحتاج أسلوب التصفح الجهد اللازم لمعرفة معاملات وآليات البحث داخل محركات البحث التي تختلف من محرك إلى آخر ؛ حيث يوفر التصفح مصطلحات البحث بدون الحاجة إلى معاملات البحث للربط بينها و صياغة إستراتيجيات البحث.

٤) يوفر التصفح المرئي لوحدات الصور الرقمية سهولة الاختيار والبحث للمستفيد من خلال التعرف إلى موضوعات الصور بطريقة أقل من تلك المستخدمة مع قوائم المصطلحات.

٥) يتميز أسلوب التصفح بالمصطلحات بأن المستفيد يستطيع من خلاله الإبحار داخل المصطلحات ذات الدلالة على محتوى و موضوع الصور الرقمية، وهو ما يوفر التعرف إلى السمات الموضوعية لوحدات الصور داخل قاعدة البينات.

٦) تستطيع المستويات المختلفة من المستفيدين التعامل بكفاية كبيرة مع أسلوب التصفح ؛ حيث إن أسلوب البحث بالكلمات المفتاحية موقوف فقط على المستفيدين أصحاب المهارات معاملات وآليات البحث على الويب.

٧) يؤدي استخدام أدوات اللغة المقيدة في تصفح المكانز وقوائم رؤوس الموضوعات إلى تحقيق درجات أكبر من التحقيق وخفض معدل الاستدعاء داخل نظم استرجاع الصور، وهو ما يعتمد على الدقة في التكشيف والبحث.

عيوب أسلوب التصفح

١) إن استخدام أسلوب التصفح بالمصطلحات يتطلب القدر الكبير من الدقة في تحليل محتوى الصور الرقمية التي تحمل معاني متباينة، وقد يؤدي تبني هذا الأسلوب إلى تشتت المستفيد في حالة الضعف الدلالي للمصطلحات.

٢) إن أسلوب التصفح المرئي لوحدات الصور يعمل بكفاية أكثر داخل قواعد البيانات الصغيرة، وهو مالا يصلح للعمل داخل قواعد بيانات الصور كبيرة الحجم؛حيث يحتاج المستفيد الكثير من الجهد والوقت لتصفح مجموعات الصور.

٣) يحتاج بناء قوائم المصطلحات الدالة داخل محركات البحث إلى إشراف العنصر البشري ؛ وهو يخالف طبيعة محركات البحث التي تعتمد على برامج ذاتية العمل قد لا ترقى إلى اعتماد المصطلحات ذات الدلالة لمحرك البحث.

٢/٢/٢/٦ استخدام البحث بالكلمات المفتاحية

يناسب هذا الأسلوب أكثر مرحلة بحث قواعد البيانات؛ حيث يجري استخدام هذه الكلمات لفتح النوافذ المختلفة لتناول الموضوع الواحد داخل قاعدة البحث. ولا يتوقف البحث بالكلمات المفتاحية عند البحث داخل حقل الواصفات الموضوعية أو رؤوس الموضوعات فقط، إنما يمكن استخدام أسماء المؤلفين أو شكل وسيط المعلومات المرئي Format مثل GIF, JPEG لتحديد وحدة واحدة من وحدات القاعدة بتمثيل مجموعة من الكلمات دون تحديد الحقول التي تنتمي إليها. وعلى هذا فإن نظام استرجاع الصور الرقمية يعمل على مضاهاة كلمات البحث مع حقول التسجيلة الببليوجرافية كاملة إضافة إلى النص الكامل لصفحة الويب المصدرية مع حقول اللغة المعيارية HTML، ليخرج نظام استرجاع الصور بوحدة/ وحدات الصور التي يتوافر في أي من حقولها كلمات البحث المدخلة إلى النظام. ولعله يمكن القول إن البحث بالكلمات المفتاحية يعد الأسلوب الأكثر شيوعا بين المستفيدين لبحث مواقع/صفحات الويب، وإن أعطى ذلك في كثير من عمليات البحث نتائج غير دقيقة خلاف التصفح بالتقسيم الموضوعي؛ ذلك لاحتياج أسلوب الكلمات المفتاحية إلى آليات بحث دقيقة يستخدمها غير المتمرسين من مستخدمي الويب. ويبرز الدور المهم لأسلوب الكلمات المفتاحية في بحث قواعد محركات الصور إذا ما تحلى نظام الاسترجاع بحلية اللغة العربية مع اعتماد اللغة الطبيعية (لغة النصوص الكاملة) كلغة تنظيم لموضوعات ووحدات الصور. ويمثل الموضوع الواحد داخل اللغة العربية أكثر من كلمة دالة (ترادف) ، كما تتخذ الكلمات المفتاحية أكثر من شكل في البحث (المفرد، المثنى، الجمع) . ويستخدم المستفيدون الأشكال المختلفة في بحث موضوع واحد وذلك بحسب طريقة التعبير عن الموضوع أو مدى تخصص المستفيدين. وعلى الجانب الآخر فإن الصور بأبعاد تحليلها الموضوعي الثلاثة يختلف التعبير عنها من شخص إلى آخر مما يولد معه واصفات متعددة للصورة الواحدة. ومن شأن أسلوب الكلمات المفتاحية تحقيق التوافق والاطراد بين الأشكال المختلفة للتعبير عن الموضوع الواحد داخل اللغة

العربية مع موضوعات الصور المتعددة والمتنوعة بصياغة إستراتيجية بحث مكونة من أشكال مختلفة للموضوع للكلمة الواحدة وموضوعات مختلفة للصورة الواحدة. ولتقريب هذا النموذج من استخدام الكلمات المفتاحية، فإن بحثا كلماته (مصر، مصريون، المصري، مصرية - الأهرام، الأهرامات، الهرم) يمكن صياغته في مصر (مع السوابق واللواحق) + and هرم (المشتقات) .

ويحتاج أسلوب الكلمات المفتاحية في بحث النصوص الكاملة بلغتها الطبيعية إلى مجموعة من آليات البحث القادرة على صياغة مختلف إستراتيجيات البحث بما يحقق الدرجة المرجوة من التحقيق في استرجاع الصور الرقمية. وتتكامل آليات بحث اللغة العربية مع آليات البحث العامة لإجراء عمليات البحث في نظام عربي لاسترجاع وسائط الصور. وتعمل الآليات العامة مثل معاملات الربط البوليني ومعاملات البحث الحر مع مختلف لغات البحث، كما تعمل آليات بحث مثل البحث بالتقابل، والبحث بالترجمة، والبحث بالمعنى، والبحث بالمشترك اللفظي... وغيرها لمعالجة خصائص اللغة العربية. ويقترح الباحث اعتماد أسلوب البحث بالكلمات المفتاحية كمحور عملية البحث والاستدعاء داخل قواعد بيانات محرك الصور مع تحقيق آليات البحث السابق ذكرها في الفصلين الثالث والرابع سواء العامة أو التي تخدم اللغة العربية، مما يمكن معه تحقيق أكبر قدر من الدقة في استرجاع الصور الرقمية العربية، ويمكن القول إن الاستخدام الأمثل لآليات بحث النظام يمكن معه تحقيق أفضل استدعاء مع أفضل تحقيق لوحدات النظام المستدعاة.

ويتميز أسلوب الكلمات المفتاحية بالكثير من المميزات التي يمكن تحقيقها مع مجموعة من العيوب التي يمكن تحويلها إلى جوانب إيجابية، قد حثت عليها دراسة[١] لإستراتيجيات بحث الويب، وهي مثل :-

فمن مميزات البحث بالكلمات المفتاحية:-

١. يمكن للمستفيد القفز السريع إلى الصفحات المحددة لخدمة الموضوع إذا ما أتقن استخدام آليات البحث.

---

(١) Keyword Searches.[with out publisher], 2002, visited 10/10/2002, http://www.concentric.net/~Rkriesel/Search/S-Keywords.htm

٢. يستطيع المستفيد الحصول على صفحات الويب غير المدرجة في قوائم البحث للأدلة إما لعدم وجود المهارة في معرفة الموضوع أو لعدم وجود الوقت الكافي للتعامل معها.

٣. يرتفع معدل حداثة صفحات الويب داخل أدوات البحث المعتمدة على برامج الزاحف والعنكبوت نظرًا لتعاملها السريع في الإضافة والمعالجة المعلوماتية أسرع من الإنسان.

وبالمقابل فإن من عيوب البحث بالكلمات المفتاحية:-

١. قد لا يستطيع المستفيد الحصول على صفحات الويب الملائمة إذا ما افتقر إلى الكلمات المفتاحية على الموضوع.

٢. لأن الكلمات المفتاحية هي العمود الفقري لمحركات البحث، فإنه لا يمكن بذلك تصفح محتويات أي من المحركات إذا ما افتقر الباحث إلى كلمات إرشادية قد يحصل عليها من خلال البناء الهرمي للموضوعات في التصفح الموضوعي.

٣. البحث بالكلمات المفتاحية داخل محركات البحث يستدعي معه الغث والثمين من المعلومات خلاف أدلة البحث التي تخضع للتقييم البشري.

٤. لا يضم محرك بحث بعينه كل مجموعات الويب، مما يستلزم معه بحث أكثر من محرك.

٥. يحتاج البحث بالكلمات المفتاحية إلى مهارة التعرف إلى آليات البحث المساعدة للكلمات المطروحة على نظام الاسترجاع .

ويتضح مما سبق أن أفضل الأساليب في استرجاع الصور هو الجمع بين الأسلوبين التصفح مع البحث بالكلمات المفتاحية.

٣/٢/٢/٦ استخدام آليات بحث اللغة العربية

تتسم اللغة العربية بخصائص فرضت على محركات البحث استخدام آليات تعامل مع المفردات العربية تعالج بها مشكلات الاسترجاع الآلي للكلمات العربية. وقد تضمن الفصل الثالث نوعي آليات البحث العربية للكلمة الواحدة وآليات البحث بكلمات متعددة. ويعد التحليل الصرفي للغة العربية باستخدام المعاجم الآلية

العامل المحوري لتفعيل آليات بحث للكلمات العربية. وعلى هذا فإن محرك بحث صور باللغة العربية يمكن بحثه بآليات البحث التالية :-

آليات البحث بالكلمة الواحدة :-

- البحث بالتطابق
- البحث غير المطابق (البحث باللواصق)
- البحث بجذر الكلمة (حل مشكلة الاشتقاق)
- البحث بالتشكيل لكلمات المشترك اللفظي
- البحث الحر
- البحث بحساسية الحالة
- البحث بالترادف
- البحث بالمتضادات
- البحث بالمعاني
- البحث بالترجمة
- إمكانية التعرف إلى الأخطاء العربية الشائعة أثناء البحث
- البحث بتشابه النطق في اللغة الإنجليزية

آليات البحث بكلمات متعددة :-

- البحث في الحقول
- البحث بالتقارب
- استخدام المنطق البوليني
- طرح سؤال للبحث: قائمة توقف عن الأحرف والأدوات
- استخدام علامات التطابق " "
- البحث المعقد أو المركب
- البحث بمثال (- /+ " " تقدير الأهمية)

ويرى الباحث أن هناك وجهين لاستخدام برنامج المحلل الصرفي؛ أولهما في مرحلة التنظيم باستخدام التحليل المعجمي لمعالجة أخطاء المكشفين عند الإدخال أو معالجة أخطاء النص الكامل للصفحات المصدرية. أما الوجه الآخر لاستخدام المحلل الصرفي، وهو ما يتناسب مع الاستدعاء لمحرك بحث الصور، أن

يعمل البرنامج على معالجة الكلمات المفتاحية المدخلة للبحث وفقًا لما أراده المستفيد من آليات بحث التضاد أو المعنى أو الترادف؛ حيث يعالج البرنامج الكلمة المدخلة في مرحلة من خلال المعاجم الآلية ثم مضاهاتها بشكلها المعالج مع كلمات النصوص والتسجيلات في مرحلة أخرى لاستدعاء الوحدات الناتجة عن المضاهاة. ويعين برنامج المحلل الصرفي على إدخال الكثير من آليات البحث الملائمة للبحث باللغة العربية مثل البحث بجذر الكلمة، البحث بالمعاني، البحث بالمقابلات في اللغات الأخرى، البحث بالمشترك اللفظي،البحث بالترادف البحث بمعالجة الأخطاء العربية الشائعة... إلخ.

٣/٢/٦ الترجيح بين النتائج (الترتيب بحسب الصلة)

يعني الترجيح بين النتائج في نظم الاسترجاع ترتيب النتائج المستدعاة بحسب الصلة Relevance مع الكلمات المفتاحية المدخلة، ويعتمد ذلك في جوهره على وزن Ranking جميع الوحدات المستدعاة مما يضع تلك الوحدات في ترتيب يساعد المستفيدين على توفير الوقت في تحديد درجة التحقيق وتقليل الجهد بإخراج أفضل النتائج. وتأتي عملية الوزن بعد حصر النظام لجميع الوحدات المطابقة لكلمات الموضوع المدخلة إلى قاعدة البيانات؛ حيث يعمل نظام الاسترجاع على المفاضلة بين الوحدات لزيادة مقدار الدقة في عرض النتائج التي يرى المستفيدون منها الجزء الأصغر فقط. ولترتيب وحدات النتائج أهمية كبيرة عند مستخدمي الويب؛ وهذا يرجع إلى زيادة أعداد الوحدات المستدعاة التي قد تصل إلى عشرات الآلاف من صفحات الويب، مما يصعب معه تصفح هذا العدد إما لقصر وقت البحث أو لما تتسم به محركات البحث من جمع الغث والثمين من مصادر المعلومات على الويب، مما يزيد بدوره من عدد وحدات الشوشرة داخل النتائج. على ذلك فان مستخدم الويب يمكن يكتفي بتصفح الوحدات الأول إذا ما اعتمد محرك البحث على آلية وزن وترتيب النتائج المسترجعة. وتزداد قيمة ترتيب نتائج البحث كلما جنح شكل الوحدات المسترجعة إلى القراءة أو كانت في شكل نصوص كاملة، وهذا يعكس معه سهولة تصفح نتائج الصور الرقمية التي تحتاج فقط إلى المشاهدة السريعة للحكم على صلة الصور المسترجعة بموضوع البحث. وبرؤية أخرى، فإن محركات الصور الرقمية تحتاج إلى آلية وزن النتائج وترتيبها نظرًا لارتفاع الأعداد المستدعاة من الصور الرقمية داخل محركات بحث الصور

والتي قد تصل إلى مئات الآلاف في بعض عمليات البحث، وعلى هذا فإن أهمية الوزن والترتيب تزداد بزيادة حجم قواعد البيانات.

وتقوم عملية وزن النتائج لتحديد مدى الصلة على عاملين أساسيين هما؛ الأول : مدى تكرار الكلمات المفتاحية المدخلة في الوحدات المسترجعة، وبحسب عدد مرات التكرار يتم ترتيب الوحدات المطابقة لكلمات البحث من الأعلى تكرارًا إلى الأقل، وبذلك يمكن تحديد أكثر الوحدات صلة بموضوع البحث. ويعتمد هذا العامل على مبدأ يشبه التكشيف الإحصائي وهو أن الكلمات الأعلى تكرارًا داخل صفحة الويب هي الأكثر صلة بموضوع تلك الصفحة. أما العامل الثاني : فهو نوع الموقع داخل صفحة الويب الذي توافرت فيه الكلمات المفتاحية المطابقة للكلمات المفتاحية المدخلة لقاعدة البيانات، وبحسب نوع الموقع يتم ترتيب النتائج. وعلى ذلك فإنه إذا ما توافرت الكلمات المفتاحية في عنوان أو مسار صفحة الويب فإن هذه الصفحة تأخذ مكانا متقدما في الترتيب من تلك التي جاءت الكلمات المفتاحية بها في حقل الوصف الموضوعي وهذه بدورها تأخذ ترتيبا في النتائج أعلى من التي جاءت بها الكلمات المفتاحية في النص الأصلي للصفحة؛ إلا أن تناول العاملين السابقين في وزن صفحات الويب وترتيب النتائج يختلف تطبيقه من محرك إلى آخر؛ حيث تفضل بعض المحركات مواقع محددة في بنية الصفحات pages structure على الأخرى، كما تأخذ بعض محركات البحث اعتبارات أخرى للحكم على مدى الصلة داخل صفحة الويب. وتعرض شركة Web Requester هذه الاختلافات في شكل ملامح مختلفة لمحركات البحث في وزن وترتيب النتائج وهي :-[١]

- إنه كلما تواردت الكلمات في العنوان كان الارتباط بالموضوع أقوى في صفحة الويب.
- الاهتمام بحقل Meta tag لما يعرض المصممون داخله من كلمات واصفة للموضوع.

---

(١) Search Engine Placement & Search Engine Ranking Service. Website Register, 2002, visited 31/10/2002,:
http://website-register.com/search_engine_ranking_standard.htm

- كثافة الكلمات المفتاحية في الصفحة؛ أي مدى توارد الكلمة في النص.
- مدى شيوع الصفحة في الصفحات الأخرى : وهو عدد روابط الصفحة المكشفة في صفحات الويب الأخرى.
- عدد الروابط غير النشطة داخل صفحة الويب مما ينقص من قيمة الترتيب.
- تقييم لغة الترميز المعيارية داخل الصفحة، وحصر أخطاء الرموز والتصميم.
- مقدار وقت التحميل الذي تقل القيمة معه كلما زاد وقت تحميل الصفحة.

ويرى الباحث أن آلية وزن النتائج بتحديد موقع الكلمات المفتاحية تلائم أكثر العمل داخل محركات بحث الصور؛ حيث تمر الصور الرقمية بأساليب تنظيم ومعالجة تختلف عن تلك المعتمدة على النصوص. ويمكن إعطاء قيم ثابتة لكل موقع داخل بنية صفحات وملفات الصور الرقمية يتم على أثرها حساب الوزن الكامل لوحدات الصور المطابقة لكلمة وعبارة البحث، وبذلك يتم ترتيب الصور المسترجعة المطابقة حسب مدى الصلة بالموضوع أمام المستفيدين مع التلميح بقيمة النتائج الأول للمستفيدين.

٣/٦ مراحل عرض النتائج

١/٣/٦ واجهة محرك بحث الصور

إن أول ما يراه المستفيدون في تعاملهم مع محركات بحث الصور هي واجهة البحث، وعلى ذلك فإن مثل هذه الواجهة تعطي الانطباع الأول عن كفاية ومجموعات نظام استرجاع الصور. وتعود أهمية واجهة البحث إلى أنها تضع المستفيد على الشكل الصحيح لاستخدام آليات بحث قاعدة بيانات الصور. وواجهة البحث ذات اللغة المعيارية تختلف عن تلكم التي تتعامل مع قاعدة البيانات، ويطلق على الواجهة الثانية اسم CGL. ولقد تبنت محركات البحث في تصميمها لواجهة البحث نوعين؛ فمنها من اختار شكل الواجهات المزدحمة بالصور مع القصور في آليات بحث المجموعات. ومنها أيضًا ما يعتمد على الواجهات البسيطة التي تهتم فقط بعرض كل آليات البحث المستخدمة في استرجاع مجموعات الصور الرقمية. ويقترح الباحث أن تكون كل آليات البحث في الواجهة الأساسية لمحرك البحث

مشتملة على كل المعاملات، وهذا يعود إلى أن السواد الأعظم من مستخدمي محركات الويب يعتمدون على إدخال الكلمات المفتاحية فور ظهور الواجهة الأساسية لمحرك البحث.

وتتمثل مواصفات واجهة البحث المقترحة في العناصر الثلاثة التالية :-

٦/٣/١/١ شكل آليات البحث

تطرح محركات بحث الصور آليات بحث في شكلين أساسين هما القوائم المنسدلة، ويعتمد هذا الشكل على تجميع كل معاملات آلية البحث الواحدة داخل قائمة واحدة؛ ذلك مثل أن يتيح محرك البحث استدعاء الصور بحسب حجم الصورة، فإن القائمة في هذه الحالة تحوي كل الأحجام التي تأتي بها الصور داخل قاعدة بيانات محرك البحث وعلى المستفيد هنا اختيار الحجم المناسب من هذه القائمة. أما الشكل الآخر من آليات بحث المجموعات فهو شكل الترميز؛ حيث توضع أمام المستفيد كل معاملات آلية البحث صفًا بجانب الاسم الدال على آلية البحث. ويتيح محرك البحث في هذه الحالة إمكانية تنشيط أو إيقاف عمل هذه المعاملات من خلال وضع علامات الترميز لكل من المعاملات المراد استخدامها في استدعاء الصور.

ويقترح الباحث استخدام الشكل الثاني لآليات البحث (الترميز)؛ لأن إتاحة مختلف معاملات البحث ظاهرة أمام المستفيد الذي يعمل على التعريف السريع بكل آليات البحث ومعاملاتها المستخدمة في صياغة إستراتيجيات بحث المجموعات، كما أنه يعطي قدرًا أكبر من الكفاية في اختيار معاملات البحث المعبرة عن الحاجة الموضوعية للمستفيد والتي يجب أن تنعكس على صياغة إستراتيجية البحث لتناسب الصور المستدعاة.

٦/٣/١/٢ عدد واجهات البحث

ينقسم المستفيدون من حيث الكفاية مع آليات ومعاملات بحث الويب إلى قسمين؛ الأول والذي لا يستخدم في بحثه سوى الكلمات المفتاحية بغض النظر عن معاملات البحث وآلياته الملائمة لاستدعاء الموضوع. أما القسم الآخر فيعتمد على صياغة إستراتيجية البحث بالشكل الذي يصلح معه استدعاء مصادر المعلومات عن

الموضوع من جانب معين. ونتيجة لما سبق فإن تعدد واجهات بحث المجموعات يشكل في ذاته الفائدة والعائق للمستفيدين وذلك بحسب مستوى التعرف إلى نظام الاستدعاء. وتضع بعض محركات بحث الصور الصفحات التي تحوي مختلف آلياتها ومعاملات بحث المجموعات تحت مسمى Advanced Search، وعادة ما يرد ذكر هذا المسمى بجوار فراغ البحث في الواجهة الأساسية، وهو ما لا يعمل عليه الكثير من المستفيدين. ومن ثم فيقترح الباحث أن تستخدم واجهة بحث واحدة على أن تتجمع بها كل آليات ومعاملات البحث التي يوفرها النظام.

٣/١/٣/٦ تقسيمات واجهة البحث

إن الشكل الكامل لإفادة مستخدمي محرك بحث الصور يعود إلى توفير مختلف أساليب وآليات البحث في واجهة واحدة، على أن تكون هذه الواجهة واضحة وبسيطة في الوقت نفسه الذي تزدحم فيه بآليات ومعاملات البحث اللازمة للتعبير عن مختلف أشكال الموضوعات. وعلى ذلك فإن التقسيم الداعم لذلك يمكن أن يكون على النحو التالي؛ أولا: أن يتم تقسيم واجهة البحث إلى قسمين أفقيًا بحيث يشمل القسم الأعلى فراغ البحث اللازم لإدخال الكلمات المفتاحية، يأتي أعلى ذلك اسم محرك البحث ثم تأتي خيارات البحث باللغة أسفل فراغ البحث بالكلمات المفتاحية. أما القسم الآخر (السفلي) من الصفحة يتم تقسيمه إلى شطرين؛ الأول إلى يمين الصفحة ويحوي هذا القسم التقسيم الموضوعي أو أسلوب البحث بالتصفح ؛ سواء اعتمد التقسيم الموضوعي هنا على الترتيب الهجائي لكل المصطلحات المستخدمة أو الاعتماد على إحدى خطط التصنيف الحصرية لموضوعات المعرفة مثل خطة ديوي أو خطة التصنيف العشري العالمي. أما الشطر الآخر من القسم الثاني (يسار الصفحة) فيضم مختلف آليات البحث المتاحة لاستدعاء الصور والتعامل مع قاعدة البيانات، بحيث توضع كل المعاملات الخاصة بكل آلية بحث بجانب اسم آلية البحث، وترتب آليات البحث بحسب أهميتها في إجراء بحوث المستفيدين. ويمكن الاستعانة بالشكل الخاص لمحركي البحث corbis, ditto لبناء الشكل المطلوب لوجهة بحث محرك الصور.

٢/٣/٦ واجهة نتائج محرك بحث الصور

١/٢/٣/٦ شكل الصور المعروضة

تحتاج محركات بحث الصور إلى صياغة ملامح أخرى غير التي تتعلق بتنظيم وبحث الصور الرقمية، وتتعلق هذه الملامح بالجانب المرئي لوسيط الصور الثابتة، ومن هذه الملامح؛ اختيار الصور الرقمية ذات الكثافة الملائمة لعرض محتوى الصورة Resolution، وأيضًا حجم اللقطات المعروضة داخل النتائج، بجانب التداخل بين البيانات واللقطات داخل وحدات النتائج المعروضة. وتأتي أهمية الشكل المعروض به النتائج لما له من تأثير على سرعة تصفح النتائج مع تحديد المستفيدين لمجموعات الدقة والشوشرة داخل النتائج. ويرى الباحث أن شكل العرض المنفرد للصور مع إضافة بعض الرسومات يمثل أفضل البدائل للتعامل مع مجموعة كبيرة من الصور التي قد تصل إلى عشرين ٢٠ لقطة في واجهة البحث الواحدة فضلا عن الرسومات التي تعطي اختيارات التعامل لكل صورة على حدة، سواء بالانتقال إلى التسجيلة الكاملة أو تحميل الصورة أو حقوق النشر... وغيرها.

٢/٢/٣/٦ البيانات الببليوجرافية المصاحبة

اعتمدت بعض محركات بحث الصور على إلحاق بعض البيانات الببليوجرافية مع الصور المعروضة في واجهات النتائج، وتؤدي الحقول الببليوجرافية مع اللقطات المعروضة إلى صغر حجم هذه اللقطات مع التداخل بين البيانات واللقطات الممثلة للصور. ويقترح الباحث عرض اللقطات منفردة مع وضع رسومات لتحديد الاختيارات والتعامل، مع حذف البيانات الببليوجرافية على أن تكون التسجيلة الببليوجرافية متاحة أمام المستفيدين دون إجراء تعديلات في البيانات المعروضة.

٣/٢/٣/٦ عدد الصور المعروضة

يرتبط عدد اللقطات أو الوحدات المعروضة في نتائج البحث مع سرعة تصفح المستفيد لهذه النتائج، على أن التحديد المسبق من جانب محرك البحث قد لا يلائم ما يرجوه المستفيد من سرعة المرور على مئات الصور الرقمية. ويرى الباحث أنه يمكن التغلب على عدد اللقطات بفتح الكثير من النوافذ أمام المستفيد

إذا ما حجب محرك البحث عدد اللقطات المعروضة، كما يفعل الكثير من مستخدمي الويب للتغلب على بطء التحميل.

٦/٣/٢/٤ الروابط داخل عرض النتائج

عملت محركات بحث الصور على إلحاق مجموعة من الروابط مع اللقطات الممثلة للوحدات المسترجعة، وتشمل هذه الروابط الانتقال إلى الصفحة المصدرية للصورة الرقمية، وأيضًا الارتباط بصفحة الصورة التي تحوي الحجم الحقيقي للصورة والذي ينتهي مساره دائما بـ GIF, JPG. وقد استبدلت الدراسة هذه الروابط بمجموعة من الرسومات النشطة المعبرة عن اختيارات التعامل مع الصورة، على أن تكون كل من هذه الرسومات رابطًا لما ترمز إليه؛ فتشير إحداها إلى التسجيلة الكاملة للصورة، أو الانتقال للصفحة المصدرية، أو تحميل الصورة، أو إرسالها كرسالة بريد إلكتروني، أو عرض الصورة فقط في الحجم الحقيقي.

٦/٤ ملامح احتياجات المستفيدين في محرك البحث

لقد أكدت دراسة C. Olivia Frost عن استرجاع مجموعات الصور الرقمية أن هناك مجموعة من الملامح التي يحتاجها مستخدمو محركات بحث الصور والتي تساعد على دعم البحث والوصول إلى الصور الرقمية داخل قاعدة البيانات. يمكن استخدام تلك الملامح لتدعيم محرك البحث المقترح هنا، ومنها[١] :-

احتياج المتمرسين من المستخدمين إلى إمكانات تقنية وفنية للتحكم وإطار الصور الرقمية وإمكانية عمل مقارنة بين ثلاث أو أربع وحدات من الصور المعروضة في شاشات النتائج مع عرض البيانات الخاصة بكل وحدة من الوحدات المعروضة.

وقد أكد المتمرسون وغير المتمرسين من المستخدمين أيضًا على إمكانية التحكم في ملفات الصور الرقمية بالنسخ أو التحميل أو الإرسال عبر البريد

(١) Frost, C. Olivia and Taylor, Bradley and Noakes, Anna. Browse and Search Patterns in a Digital Image Database, University of Michigan, 2002?, cuted at 6/10/2003, visited at http://www.si.umich.edu/~cfrost/papers/fidel.html

الإلكتروني؛ إلا أن ذلك يعتمد بقدر كبير على حقوق النشر والاستخدام المتعلقة بالصور على الويب.

الاحتياج إلى إجراءات أخرى للمساعدة في البحث عن الصور مثل تهذيب نتائج البحث بتوفير إمكانية البحث داخل النتائج، أيضًا إمكانية البحث بأسماء المصورين "المؤلف"، كما يمكن استرجاع الوحدات المشتركة مع وحدة من النتائج، إذا ما وجد المستفيد بها الخصائص الملائمة لموضوع البحث أسفل عبارة " صور أخرى مشابهة لهذه " .

أن يتمتع عرض الصور داخل محرك البحث بكثافة عالية للصور المعروضة حيث يمكن اعتمادًا على ذلك زيادة عدد الوحدات في شاشات النتائج الواحدة، أو الاستفادة من ذلك عند تحميل تلك الصور.

٥/٦ الخلاصة

عملت الدراسة في هذا الفصل على تبني مجموعة العناصر المقترحة لبناء محرك بحث عربي يدعم خصائص اللغة العربية. وقد بدأت هذه العناصر بالحديث عن ملامح محددة لتكشيف ملفات الصور الرقمية على الويب والتعامل معها من ناحية تنظيم وحدات المعلومات على الويب. ثم الجزء الآخر عن الجوانب العملية لمحرك بحث ربي للصور الرقمية، بداية من اقتناء ملفات الصور ثم مرحلة التكشيف والبحث لوحدات الصور الرقمية، إلى جانب الحديث عن استخدام اللغة العربية داخل محرك البحث والعمل بها في عمليات التنظيم والبحث.





نتائج الدراسة

نتائج الدراسة

توصل الباحث إلى مجموعة من النتائج يجملها فيما يلي:

١. إن التعدد والتنوع في أشكال وسائط المعلومات الرقمية المرئية التي أفرزتها بيئة الويب، يحتاج بدوره إلى تكثيف الدراسات داخل علم المكتبات والمعلومات للخروج بالأطر والمعايير التي يمكن بها بناء مواصفات تنظيم واستدعاء لمثل هذا الخضم الهائل المتنوع من أوعية المعلومات. ويرتبط بذلك أن المفهوم العام للإنترنت إنما هو مصدر معلومات شامل لكافة المعلومات بمختلف أشكالها.

٢. إن التنامي السريع لمحتوى الإنترنت قد أثر على ملاحقة تطور وتعدد وسائطه النصية والصوتية والمرئية، بما يعكس عدم كفاية هيئة واحدة لرعاية تنظيم المعلومات على الإنترنت.

٣. إن أساليب بحث الويب وتنوع أدوات البحث على الويب مازالت في حاجة أكبر إلى تطويرها، حيث إن أغلب أدوات البحث تعمل بما يوافق المستوى الأدنى من مهارات المستفيدين في بحث معلومات الويب. وهو ما يؤثر بدوره على تطوير آليات بحث مصادر الويب غير النصية .

٤. تتواجد نظم استرجاع الصور في مستويين مختلفين؛ الأول: نظم إدارة الصور الرقمية وهي نظم تعمل على المعالجة الفنية الكاملة لملفات الصور من تنظيم ومعالجة ثم توفير أساليب البحث الملائمة لاسترجاع ملفات الصور. الثاني: نظم بحث وتصفح الصور ومنها محركات بحث الصور وهي نظم تعمل على بحث تصفح الصور من خلال النصوص المحيطة بها دون تحليل ملفات الصور.

٥. تركز محركات البحث غير العربية على ملفات بعينها في إدخال ملفات الصور الرقمية داخل قواعد بياناتها، وهو ما يساعد على تشكيل الصورة الذهنية أو تحويل اتجاه الرأي العام لمستفيدي الويب من خدمات بحث الصور في بعض القضايا العامة السياسية أو الاجتماعية، وذلك في استخدام بعض الكلمات البحثية التي تشكل بذاتها قضايا عالمية مثل العراق، الشرق الأوسط، الإرهاب، الإسلام ... إلخ، مما يجعل من محركات بحث الصور الرقمية الثابتة تؤدي دروا غير بريئ في التعبير عن قضايا محددة .

٦. جاءت نسب تحقيق محركات بحث الصور على الويب ضعيفة للغاية، ويعود السبب في ذلك إلى غياب معيار ثابت لبناء ملفات للصور الرقمية وبداخلها التسجيلة الببليوجرافية الواصفة مع مجموعة الكلمات الواصفة للعناصر الموضوعية داخل الصورة سبب ضعفا في بناء ملفات الصور الرقمية .

٧. إن بناء قواعد بيانات الصور الرقمية داخل الكثير من المؤسسات مثل المتاحف وهيئات الأبحاث الفضائية، بل والأرشيفات الصحفية لن يتم بشكل صحيح إلا بالاعتماد على عاملين هما ؛ تبني معيار لوصف الصور الرقمية مثل دبلن كور مع إجراء التعديلات اللازمة عليه، إلى جانب استخدام تقنيات بناء ملفات الصور الرقمية التي تضم النص إلى جانب المحتوى المرئي في ملف واحد لا ينفصل به النص عن الصورة.

٨. يعد معيار دبلن كور أفضل معايير الوصف الفني التي تتعامل مع الصور الرقمية؛ حيث يعتبر هذا المعيار أكثر المعايير اهتماما بوصف المحتوى من خلال تخصيص عنصري الموضوع subject والوصف description إلى جانب عناصر الوصف المادي، كما أن هذا المعيار يسير في تطوير مستمر بما يمكن له معالجة ملفات الصور الرقمية بلغة XML.

٩. أوضحت دراسة استرجاع الصور الرقمية العزلة التامة في العمل بين استرجاع الصور المبني على النص واسترجاع الصور المبني على المحتوى. ويتبع الاسترجاع بالمحتوى تقنيات تحليل الصورة باستخدام البنية والشكل واللون، خلاف ما يعمل به الاسترجاع بالنص في الاعتماد على التكشيف والبحث للكلمات المفتاحية. وأن الجمع بين النوعين في نظام واحد يفيد باحثي الصور

في استرجاع صور ذات بنية أو شكل أو لون محدد بنفس الحاجة الموضوعية للباحث.

١٠. تعمل جميع محركات البحث باستخدام آلية واحدة لتنظيم وبحث مصادر الويب، وذلك على الرغم من اختلاف وسائط حمل المعلومات وتعددها داخل الوسيط الواحد. وأكد ذلك التوحد في بناء محركات البحث اعتمادها على ثلاثة برامج فرعية لجمع وتنظيم وبحث معلومات الويب المختلفة، تعمل هذه البرامج بآليات مخصصة لاسترجاع النصوص.

١١. تعتمد محركات الويب عامة ومحركات بحث الصور خاصة على برامج الزاحف في انتقاء وإضافة مصادرها من أوعية المعلومات، وعلى ذلك فإن الفلاتر (المرشحات) داخل أدوات بحث الويب تعمل من خلال بناء سياسة التوقف والمرور Stop / Go list داخل برنامج الزاحف.

١٢. إن تنظيم وبحث المادة المصورة الرقمية بصفة عامة والمصورة الثابتة بصفة خاصة، يعتمد على التحليل الموضوعي لهذه الوسائط أكثر من اعتماد محركات البحث على السمات المادية أو استخدام تقنيات المطابقة للألوان أو التركيب، إلى جانب احتياج تنظيم المادة المصورة للتطوير الحتمي لبرامج الزاحف التي تتعامل مع النص المصاحب فقط لملف الصورة.

١٣. إن تنظيم ومعالجة ملفات الصور الرقمية يمكن أن يتعدى عناصر الوصف الفني إلى تقنيات تحليل ألوان وبنية ملفات الصور الرقمية. وتسهم المعالجة الرقمية إلى جانب المعالجة الفنية للصورة في رفع درجات التحقيق باسترجاع صور ذات بنية وشكل محدد لنفس المحتوى الموضوعي. وبذلك يمكن الجمع بين الاسترجاع المبنى على المحتوى والاسترجاع المبني على النص.

١٤. تحتاج المعالجة الفنية للصور الرقمية إلى العنصر البشري أكثر من برامج محركات البحث. ويؤكد ذلك أن الاطراد بين الكلمات الوصفية والكلمات البحثية أيسر في جانب الوصف المادي، خلاف ما يحدث من التفاوت بين الكلمات الوصفية والكلمات البحثية في تحليل المحتوى الموضوعي. ويأتي ذلك من اعتماد برامج محركات البحث على أسلوب الاقتباس للنصوص المحيطة بالصورة أكثر من أسلوب التحليل لمحتواها .

١٥. يعتمد استخدام أسلوب البحث الحر أو أسلوب التصفح على طبيعة أداة البحث في الويب، حيث تعتمد محركات البحث في حالة المادة النصية على استخدام البحث الحر أكثر من محركات البحث الخاصة بالمادة غير النصية. وأن الجمع بين الأسلوبين أفضل في البحث مع محركات بحث الصور.

١٦. إن الاعتماد على اللغة المقيدة في بناء أسلوب التصفح يمكن محرك بحث الصور من تحقيق أعلى درجات الاطراد بين تحليل محتوى الصور وبين استدعائها. وذلك بالاعتماد على أحد المكانز أو إحدى قوائم رؤوس الموضوعات. وتتيح إمكانية تصفح هذه المكانز داخل محرك البحث تعرف المستفيدين على موضوعات الصور، إلى جانب فهم كل موضوع على حدة في إطار السياق العام.

١٧. تعد آلية البحث بالعبارات هي أفضل آليات البحث المستخدمة بين محركات بحث الصور، يلي ذلك آلية البحث بالمنطق البوليني. ويرجع ذلك إلى ملاءمة البحث بالعبارات لسهولة استخدامها بين مستخدمي محركات البحث بصفة عامة، بالمقارنة مع آلية المنطق البوليني الذي يحتاج إلى عقلية رياضية.

١٨. تحتاج آلية البحث بحقول الوصف في محركات البحث إلى معايير ثابتة للتعامل مع وصف الصور الرقمية؛ حيث استخدمت محركات البحث آلية البحث بالحقول على استحياء من عدم وجود تسجيلات نهائية لملفات الصور تجمع بها حقول الوصف، جاء ذلك من خلال وضع بعض محددات لبحث الصور كالاختيار بين نوع الصورة و شكل ملف الصورة ولون الصورة و الحجم...الخ.

١٩. إن اللغة العربية وإن كانت قد حظيت بمعايير البحث الآلي داخل قواعد بيانات الويب العربية أو التي تدعمها، فإنها لم تحظ بهذا الاهتمام في محركات بحث الويب وخاصة تلك التي تتبنى البحث فقط في ملفات الصور الرقمية، وهذا إنما يعود لندرة أو انتفاء وجود محركات بحث صور عربية أو غياب ملفات الصور الرقمية الثابتة عن ذهن الباحثين في تنظيم المعلومات العربية.

٢٠. إن محركات بحث الويب عامة ومحركات بحث الصور خاصة الداعمة للغة العربية إنما تعتمد على تحقيق استدعاء مرتفع مع درجة تحقيق منخفضة في

نتائج البحث أمام مستخدم الويب. حيث تغلب محركات البحث عامل الاستدعاء على عامل التحقيق في استرجاع نتائج كثيرة العدد لا تفي بالتخصيص المطلوب لمستخدم الويب. وهذا يرجع إلى ضعف الخصائص الآلية لمحركات البحث في دعمها اللغة العربية.

٢١. إن مشكلات التعامل مع الصور الرقمية في البيئة الافتراضية العربية تتعدى حدود المشكلات اللغوية الصرفية إلى حد مشكلات المعنى والثقافة. وتستخدم تقنية التحليل الصرفي للتعامل مع المشكلات الصرفية؛ إلا أن مشكلات المعنى والثقافة تحتاج إلى إشراك العنصر البشري في التعامل مع محتوى الصورة.

٢٢. أظهر اختبار آليات بحث اللغة العربية مدى افتقار محركات البحث في استخدام تقنيات التعامل مع اللغة العربية. وهذا يرجع تحديدًا إلى قلة اهتمام محركات بحث الصور في ضم ملفات الصور العربية نسبة إلى تلك الإنجليزية. ويعد محرك البحث Google أفضل محركات الدراسة في التعامل مع اللغة العربية، غير أنه لم يستخدم أيًّا من آليات البحث المتخصصة للتعامل مع اللغة العربية.

٢٣. أوضحت دراسة حقول الوصف الببليوجرافي داخل محركات البحث أن محركات بحث الصور لا تعمل من خلال معايير محددة، إنما تعمل اجتهادا لتوفير أكثر البيانات الببليوجرافية أهمية بالنسبة لباحثي الصور الرقمية. وقد كانت محركات البحث Corbis, Picsearch, Altavista, Fast على الترتيب أفضل محركات الدراسة في التعامل مع حقول الوصف الببليوجرافي ؛ حيث عمل Corbis على توفير أغلب حقول الوصف للبحث أمام المستفيد. أما Picsearch فقد قدم أفضل أشكال التسجيلات النهائية لملفات الصور الرقمية. وجاء بعده محركا Altavista و Fast في صنع التسجيلة النهائية.

٢٤. اتفقت جميع محركات البحث في العمل على استخدام نص صفحات الويب بوصفه مصدرًا أوليًّا لتكشيف وتحليل ملفات الصور الرقمية التي جاءت بداخلها. وهو ما يعني أن التكشيف بالاشتقاق لصفحات الويب مازال أساس العمل داخل محركات البحث. وقد شذ عن ذلك فقط محرك البحث Corbis؛

حيث أثر التدخل البشري على اختيار الكلمات المفتاحية بإدخال التحليل البشري لموضوعات الصور الرقمية.

٢٥. انقسمت البيانات التي تصاحب عرض الصور إلى قسمين ؛ الأول: البيانات الوصفية التي تعطي وصفًا لحجم وشكل ونوع الصورة وما إلى ذلك من العنوان والمسار. والثاني : خدمات التعامل مع الصورة ؛ حيث تعرض هذه البيانات الخيارات المطروحة أمام المستفيد للتعامل مع الصور المسترجعة. وقد تقدم محرك البحث Google على كل المحركات الأخرى في عرض أكبر قدر من البيانات المصاحبة للنتائج ثم تساوت جميع المحركات بعد ذلك في عرض البيانات المحدودة. ويعد محرك البحث Corbis أفقر محركات البحث في القسم الأول وأغناها في طرح اختيارات التعامل مع ملفات الصور؛ حيث يعرض إمكانية التحميل أو الشراء أو الإرسال.

٢٦. نظرًا لأن محركات بحث الصور تعمل بطريقة تفاعلية غير ثابتة (ديناميكية) ، فإنه لا يصلح تطبيق أساليب قياس الكفاية في نظم الاسترجاع الثابتة على محركات البحث. ومن ثم فإن نظام لانكستر المخصص لأساليب قياس الكفاية في النظم الثابتة لا يصلح مع النظم غير الثابتة أو بمعنى محركات الويب.



التوصيات

توصل الباحث إلى مجموعة من التوصيات يوجزها في الفئات التالية:

الفئة الأولى: المكتبات الافتراضية

١. يجب على المكتبات العربية بوصفها مؤسسات خدمات المعلومات السعي لبناء مجموعتها من الصور الرقمية كأحد وسائط المعلومات المتعددة، وذلك لما يحمله وسيط الصورة الآن من أهمية في بحث المستفيدين في ظل قضايا تتعامل أكثر مع الجوانب الإعلامية والرؤية للأحداث أكثر من وصفها بالكلمات والألفاظ.

٢. يجب على الشركات العربية العاملة في مجال الويب والداعمة لمحركات البحث الاتجاه نحو بناء الكثير من محركات البحث الداعمة للبحث في تنظيم وبحث المصادر العربية على الويب ومنها ملفات الصور الرقمية.

٣. يجب على المؤسسات العربية ولاسيما الإعلامية التي تتعامل مع ملفات الصور على الويب، الوعي الكامل بمفاهيم التكشيف في البيئة الافتراضية. فالبحث بالصور لا يعني في ذاته فقط تقديم خدمة بحث ملفات الصور الرقمية الثابتة بجانب باقي مصادر الويب، إنما هو محاولة للتعبير عن بعض اتجاهات الصور العربية في القضايا التي يناقشها العالم مثل قضايا الشرق الأوسط والإرهاب والإسلام والقضية الفلسطينية. ويتجلى من ذلك الدور الإعلامي الذي يمكن أن تؤديه محركات بحث الصور الإعلامية، بل يجب بناء محركات بحث صور ثابتة تدعم فقط الصور الرقمية الإعلامية.

الفئة الثانية: أقسام المكتبات والمعلومات ومراكز المعلومات

٤. يجب على أقسام المكتبات والمعلومات في الوطن العربي الاهتمام بدراسات عن نظم الاسترجاع في البيئة الافتراضية وذلك من خلال المناهج الدراسية

أو الدراسات العليا لما له من أثر بعد ذلك في أن تكون الويب أحد مصادر المعلومات المرجعية المهمة في خدمات المكتبات.

٥. يجب على أقسام المكتبات والمعلومات في الوطن العربي أخذ المبادرة بتأصيل مفاهيم تنظيم وبحث الأشكال المختلفة لوسائط المعلومات على الويب للتمييز بينه وبين الأبحاث التي تحيد نحو علوم الحاسبات أو الجوانب الفنية.

الفئة الثالثة: المواصفات والمعايير الدولية

٦. يجب على الشركات المنتجة والداعمة لملفات الصور الرقمية تبني معيار وصف للصور لإنتاج ملفات الصور بالشكل الذي يتيح لمحركات البحث تكوين مجموعاتها من الصور بشكل يسهل عملية البحث. ويقترح أن يكون معيار دبلن كور هو البداية باعتباره أفضل المعايير حتى الآن إلى أن يتم التوصل إلى معايير متخصصة في إنتاج ملفات الصور.

٧. يجب على الجمعيات العربية التي تعمل على تعريب بنية الويب الاعتماد على مجموعة من القواعد والمعايير الموحدة لمحركات البحث العربية، مع العمل المشترك لإنتاج ميتاداتا عربية تساهم في تجنب أخطاء تنظيم وبحث مصادر معلومات الويب العربية.

٨. يجب على المتخصصين في علم المعلومات وعلم اللغة دعم اللغة العربية كلغة محتوى وتنظيم لأوعية المعلومات بالمزيد من التطوير في معايير وتقنيات التخزين والاسترجاع الآلية لما تتسم به اللغة العربية من سمات عدة كالترادف والاشتراك اللفظي واللواصق ... إلخ التي تختلف بها عن غيرها من اللغات.

وفي ظل هذه التوصيات يقترح الباحث مجموعة من الدراسات المقترحة لباحثين آخرين :-

- المادة الصوتية على الويب : دراسة لتنظيم وبحث الملفات الصوتية.
- محركات بحث اللقطات المتحركة "الفيديو" على الويب، لما تتسم به اللقطات من طبيعة تكشيف تختلف عن الأفلام.
- دراسة تحليلية متعددة الجوانب للمؤسسات الإعلامية على الويب في تنظيم واسترجاع الصور الإعلامية الرقمية.

المراجع

أولا: المراجع العربية الورقية

١. خالد محمد رياض.

أدلة ومحركات بحث شبكة الإنترنت : دراسة مقارنة.- الإستراتيجية العربية الموحدة للمعلومات في عصر الإنترنت ودراسات أخرى، الاتحاد العربي للمكتبات والمعلومات، تونس، (أكتوبر١٩٩٩).- ١٣٩ ص.

٢. زين عبد الهادي.

محركات البحث على شبكة الإنترنت : دراسة تجريبية مقارنة، مجلة المكتبات والمعلومات العربية، ع٢، أبريل ٢٠٠٢.ص ١٠.

٣. شعبان عبد العزيز خليفة و محمد عوض العايدي.

المواد السمعية و البصرية و المصغرات الفيلمية في المكتبات و مراكز المعلومات.- ط٢، مزيدة و منقحة. [القاهرة]: مركز الكتاب للنشر، ١٩٩٧م. ٢٣٠ص.

٤. علي السليمان الصوينع.

استرجاع المعلومات في اللغة العربية.- الرياض: مكتبة الملك فهد الوطنية، ١٩٩٤. ١٨٠ص.

٥. غولد، تشيرل.

البحث الذكي في الإنترنت: أدوات وتقنيات للحصول على أفضل النتائج؛ ترجمة عبد المجيد أبوعزة.- الرياض: مكتبة الملك فهد الوطنية، ٢٠٠١م ١٤٨, ص.

٦. لانكستر، ولفرد.

نظم استرجاع المعلومات؛ ترجمة حشمت قاسم._ القاهرة :دار غريب للطباعة والنشر والتوزيع، ١٩٨١. ١٦٠ ص.

٧. هانئ محيى الدين عطية.

جهود تصنيف المعرفة في البيئة الإلكترونية : دراسة استكشافية.- الأكاديمية للمكتبات والوثائق والنظم المعلوماتية، ع١ (٢٠٠٤).- ص٥-٧٧.

٨. هانئ محيى الدين عطية.

تكشيف الصور بين العلم والتعلم : ورقة في ملف القضية الفلسطينية.- أسيوط: المؤتمر القومي الخامس لاختصاصي المكتبات والمعلومات والأرشيف في مصر، ٢٠٠١م.

ثانيًا: المراجع العربية الإلكترونية

٩. أبو السعود إبراهيم.

المحور الرابع : التجارب العربية الرائدة في مجال حفظ و تنظيم الوثائق والتأهيل و التدريب الوثائق التاريخية أهميتها ووسائل الحفاظ عليها وتجربة الأهرام في تنظيم وحفظ الوثائق وإتاحتها للمستفيدين، النادي العربي للمعلومات، ٢٠٠١م

http://www.arabcin.net/arabic7/nadweh/print_pages/print_%20fourth_pivot/historical_doc.htm

١٠. أحمد محمد المعتوق.

الألفاظ المشتركة المعاني في اللغة العربية طبيعتها ـــ أهميتها ـــ مصادرها، جامعة الملك فهد للبترول والمعادن، تاريخ الاستشهاد ٢٠٠٢/١٠/٢٢م

http://www.uqu.edu.sa/majalat:/shariaramag/mag21/MG-012.htm

١١. إرشادات البحث في الإدريسي. ٢٠٠٢/١٠/١٠

http://asb.kacst.edu.sa/IDRADMIN/Arabic/Help/AT:ips.htm

١٢. عبد الرحمن بن حسن المحسني.

قضية الترادف : النظرية والتطبيق، تاريخ الاستشهاد ٢٠٠٢/١٠/٢٣م

http://www.adabiabha.com/biader/4.htm

١٣. الشوكاني.

مباحث لغوية في كتب الأصول : إرشاد الفحول إلى علم الأصول. ٢٢/١٠/ ٢٠٠٢

http://www.angelfire.com/tx4/lisan/mawdouat:.htm:

١٤. مساعد بن صالح الطيار.

محركات البحث العربية وطرائق البحث فيها، الجزيرة، ٢٠٠١، ١٠/١٠/٢٠٠٢م

http://www.al-jazirah.com.sa/evillage/22112001/wr1.htm

١٥. وكالة الأنباء الكويتية كونا.

البحث في قاعدة البيانات، الوكالة، ٢٠٠٠م، ١٥/١٠/٢٠٠٢م

http://www.kuna.net.kw/InfoCenter.asp

ثالثًا: المصادر الأجنبية الإلكترونية

1. A Metadat:a Framework to Support the Digital Resource Management. Texas university, 2003, 10/10/2003. : http://texashistory.unt.edu/guides/Draft-UNTL-Metadat:a-Guideline.htm.

2. AL-BADR, Badr H. Using the Internet in Arabic: Problems and Solutions, King Abdul aziz City for Science and Technology, [2002?], visited 22/11/2002, : http://www.isoc.org/inet98/proceedings/5f/5f_1.htm

3. Al-Kharashi, Ibrahim A. A Web Search Engine for Indexing, Searching and Publishing Arabic Bibliographic Dat:abases, King Abdul Aziz City for Science and Technology, Computer and Electronics Research Institute, [2001?] 31/12/2002, visited at: http://www.isoc.org/inet99/proceedings/posters/085/

4. Barlow, Linda. A Helpful Guide to Web Search Engines, Monash Informat:ion Services, 2002, 11/112002, visited at: http://www.monash.com/spidap.html

5. Beekink, Macel.Using scientific internet subject directories, training end users. (2000), (19/10/2001), visited at: http://southernlibrarianship.icaap.org/content/v03n03/Hoggan_d01.htm

6. Berinstein, paula. Turning visual : image search engines on the web, on line inc., 1998 (2003), :http://onlineinc.com/onlinemag/OL1998/berinstein5/html

7. Callery, Anne. Yahoo! Cataloging the Web, Yahoo! Inc., 1996, visited at: 20/10/200, http://www.library.ucsb.edu/untangle/callery.html

8. Ciolek, T. Matthew. Internet Structure and Development: On Strat:egic Uses of the Archetypes of the Networked Mind, author, 1999, visited at: 10/10/2003, : http://www.ciolek.com/PAPERS/pnc-taipei-99.html

9. Clyde, Anne. 2000. Search engines : an overview, teacher librarian, visited at:18/9/2002, : http://www.hi.is/~anne/websearch_bibliography.html

10. Cohen, Laura. Conducting Research on the Internet, university at: Albany libraries, 2003, visited at: 1/6/2003, : www.uah.edu/502/Elmore_POTW_Ch4_report.

11. Controlling your language - links to metadat:a vocabularies. TASI Technical Advisory Service for Images, 2003, visited at: 12/12/2003, : http://www.tasi.ac.uk/resources/vocabs.html

12. Corfman, Robert. Satisfying the Search Engines - Key Words Meta Tag, the author,1998, visited 6/2/2003, : http://www.corfman.com/robert/CityU/SearchEngine/keywords.htm

13. Cristine, Jake. stat: of the arts : current applications for indexing images, the author, 1999, visited at: 23/9/2002, : http://portal.acm.org/citat:ion.cfm?id=614788

14. Day,Michael. Metadat:a for images: emerging practice and standards, UKOLN: The UK Office for Library and Informat:ion Networking, 1999, visited at: 12/2/2004, : http://www.ukoln.ac.uk/metadat:a/presentat:ions/cir99/paper.htm

15. Denenberg, Ray. Structuring and Indexing the Internet, Library of Congress, 1996, visited 1/11/2003, : http://lcweb.loc.gov/z3950/agency/papers/italy.html

16. DIG35: Metadata Standard for Digital Images. XMLCoverPages, 2002, visited at: 21/6/2004, : http://xml.coverpages.org/dig35.html

17. DIMP: The Use of Images to Support Instruction and Presentation. TASI Technical Advisory Service for Images, 2003, visited at: 11/12/2003, : http://www.tasi.ac.uk/advice/using/ppt_case.html

18. Drabenstot, Roren M.. Web search strategy development,gale group, (2001), visited (2/4/2002), : http://www.sciencemag.org/cgi/content/abstract

19. Eakins, John P and Graham, Margaret E. Content-based Image Retrieval A report to the JISC Technology Applications Program, the institute, 1999, visited at: 24/9/2002, : http://portal.acm.org/citat:ion.cfm?id=614839

20. Eriksson Yvone. How to make tactile pictures understandable to blind reader, internat:ional federat:ion of library associat:ions and institutions(IFLA), 1999, visited (1/12/2001), : http://www.tasi.ac.uk/advice/creat:ing/fformat:.html

21. Farace, Joe,. Take Control of Your Image Assets, computer user, visited at: 21/9/2002, 2000,: http://www.computeruser.com/articles/1907,2,3,1,0701,00.html

22. File Format:s. TASI Technical Advisory Service for Images, 2003, VISITED AT: 24/4/2003, : http://www.tasi.ac.uk/advice/creat:ing/fformat:.html

23. Forger, Garry. Image, University of Arizona Library, 2002 ?, visited at: 5/10/2003, : http://dizzy.library.arizona.edu/sitesrch/train.html

24. Frequently Asked Questions - Picture Search. Fast Search & Transfer ASA, 2002, visited 20/12/2002, : http://www.victoria.tc.ca/Resources/www.services.html

25. Frequently Asked Questions. Web Drive Canada Inc., 2001, visited 30/1/2003, : http://www.webdrive.ca/faq.html

26. Frost, C. Olivia and Noakes, Anna. Browsing Images Using Broad Classificat:ion Cat:egories, the University of Michigan, [2000?], visited 1/6/2002, ited at: http://www.si.umich.edu/~cfrost/

27. Frost, C. Olivia and Taylor, Bradley and Noakes, Anna. Browse and Search Patterns in a Digital Image Database, University of Michigan, 2002?, cuted at: 6/10/2003, : http://www.si.umich.edu/~cfrost/papers/fidel.html

28. Goodrum, Abby A.. Image information retrieval: an overview of current research, informing science, 2000, visited at: 1/10/2003, : http://64.233.161.104/search?q=cache:BRhAug3cfd

29. Graham, Margaret E. THE DESCRIPTION AND INDEXING OF IMAGES: Report of a survey of ARLIS members, Institute for IMAGE DAT:A Research1998/99, visited at: 6/10/2003, : http://www.unn.ac.uk/iidr/ARLIS/.

30. How Search Engines Work. Search Engine Wat:ch, 2002, visited at: 18/4/2003, : http://www.searchenginewat:ch.com/webmasters/article.php/2168031

31. Imad Hamdan. Web Site Promotion Techniques for Top Search Engine Listings The Internet Marketing Warriors, 2003, visited at: 11/9/2003: http://www.vastdream.com/TopListingsReport2.html

32. Image Management Systems. Technical Advisory Service for Images, 2003, visited at: 20/10/2003: http://www.tasi.ac.uk/advice/delivering/ims.html

33. Image-Finding Strat:egies. Washington University, 2002, visited 30/11/2002: http://library.wustl.edu/subjects/art/im

34. Image Tagging: Embedding Metadata In Image Files. TASI Technical Advisory Service for Images, 2003, visited at: 28/4/2003: http://www.tasi.ac.uk/advice/delivering/faqtagging.html

35. Karnowski, T.P. image retieval, Image Science and Machine Vision, 2002, visited at: 12/5/2004: http://www-ismv.ic.ornl.gov/projects/video.htm

36. Lazewski, Barbara. WWW Search Strat:egies, University of Wisconsin-Madison, 2002, visited 30/11/2002: http://www.library.wisc.edu/libraries/Steenbock/services/wwwstrat:.htm

37. Lu, Guojun and Williams, Ben. An Integrat:ed WWW Image Retrieval System, the author, 1999, visited at: 6/10/2002: http://ausweb.scu.edu.au/aw99/papers/lu/paper.html

38. Martinez, anne GIF vs. JPG vs. PNG. mcgrawhill companies, inc,2001, citd (20/7/2002): http://www.cheapwebrich's.com/articles/gif_jpg_png.html

39. Martinez, Aanne. GIF vs. JPG vs. PNG. Mc GrawHill companies, 2001 inc, citd (20/7/2002): http://www.cheapwebrich's.com/articles/gif_jpg_png.html

40. Mat:tison, david. Images of History on the Web. informat:ion today,inc, 2002, visited at: 15/10/2003: http://www.infotoday.com/searcher/may02/mattison.htm

41. Metadat:a: Image Description, Indexing and Cat:aloguing. tasi, 2004, visited at: 21/6/2004, : http://www.tasi.ac.uk/advice/delivering/meta.html

42. Miller, Steven J. Metadat:a and Cat:aloging Online Resources, University of Wisconsin–Milwaukee Libraries, 2003, visited at: 10/5/2004, : http://www.uwm.edu/~mll/resource.html

43. Notess, Greg. Searching beyond text : issues with multimedia searching, informa-t:ion today, 2000, visited 27/9/2002, visited at: http://www.onlinemag.net/OL2000/net9.html

44. Pedley, Paul. The Invisible Web, The Library Associat:ion, 2000, visited at: 20/10/ 2003, : http://www.la-hq.org.uk/directory/record/r200011/article2.html

45. PETERSON, RICHARD EINER. Eight Internet search engines compared, first Monday, 1997, visited at: 21/6/2003, : http://www.firstmonday.dk/issues/issue2_2/peterson/

46. Reichmann, Michael H. Understanding resolution, luminous landscape, 2002, visited20/12/2002, : http://www.luminouslandscape.com/understanding resolution.htm cite

47. Report of a survey of ARLIS members, Institute for IMAGE DAT:A Research1998/99, visited at: 6/10/2003, : http://www.unn.ac.uk/iidr/ARLIS/.

48. Rusch-Feja, Diann. Metadat:a: Standards for Retrieving WWW Documents (and Other Digitized and Non-Digitized Resources), Astronomical Society of the Pacific, 1998, visited at: 2003: http://www.eso.org/gen-fac/libraries/lisa3/ruschfejad.html

49. Search Engine Placement & Search Engine Ranking Service. Website Register, 2002, visited 31/10/2002: http://websiteregister.com/search_engine_ranking_standard.htm

50. Search engines strat:egies guide : spider indexed search engines. search engine strat:egies, [2001],visited 5/1/2003: http://www.searchenginestrat:egies.biz/spi-der-indexed-search-enginedirectories.htm

51. Searching for and Retrieving Digital Images. TASI Technical Advisory Service for Images, 2003, visited at: 25/4/2003: http://www.tasi.ac.uk/advice/using/image-use.html

52. searching the internet :recommended sites and search techniques. on line proceed-ings, 6/2002: http:/library.albany.edu/internet/search.html

53. Shirley Aw. How to Search Medline on OVID: Boolean Operat:ors and Truncat:ion, The Nat:ional University of Singapore Library, Jan 2000,: http://www.netmechanic.com/news/vol5/promo_no3.htm

54. Smith, John R. & Chang, Shih-Fu. Searching for Images and Videos on the World-Wide Web, Department of Electrical Engineering and Center for Image Technology

for New Media, Columbia University, New York, 1996, visited at: 12/12/2002: http://www.dlib.org/dlib/february97/columbia/02chang.html

55. Subject Tree Searches : Intro To Subject Trees.[w. p.], 2003?, visited 28/8/2003: http://www.concentric.net/~Rkriesel/Search/S-Trees.html

56. Sullivan, Danny. Search Engine Optimizat:ion & Marketing Glossary, Submitawebsite, Inc, 2004, visited at: 12/6/2004: http://www.submitawebsite.com/seo_glossary.html

57. Sullivan, Danny. Deat:h Of A Meta Tag, Jupiter media Corporat:ion, 2002, visited 25/12/1/2002 : http://searchenginewat:ch.com/subscribers/articles/02/10-meta.html

58. Taylor, Chris. metadat:a, University of Queensland Library, 2003, visited at: 10/5/2004: http://www.cybrary.uq.edu.au/iad/mainmenu.html

59. Thomason, Larisa. Promotion Tip: Search Engines Get The Picture, 2002, visited 25/12/2002: http://www.netmechanic.com/news/vol5/promo_no3.htm

60. Uses for Image Applicat:ions. Hewlett-Packard Development Company, 2005, visited at: 23/5/2005, visited at: http://docs.hp.com/en/B2355-90068/ch01s01.html

61. Wang, James Ze...[etc]. Semantics-sensitive Retrieval for Digital Picture Libraries, D-Lib Magazine, 1999, visited at: 12/10/2003: http://www.dlib.org/dlib/november99/wang/11wang.html

62. WEEKLY SEARCH ENGINE FACTS. The Axandra newsletter archive, 2003, visited at: 1/10/2003: http://www.axandra.com/news/index.htm

63. search engines - submissions advice. w.p., 2003,visited at: 21/6/2003: http://alt.venus.co.uk/weed/search/advice.htm

64. About NISO, niso, 2001, visited at: 20/9/2002: http://www.niso.org/about/index.html

65. Keyword Searches.[with out publisher], 2002, visited 10/10/2002: http://www.concentric.net/~Rkriesel/Search/S-Keywords.htm

66. Definitions of News group on the Web.google, visited at: 12/6/2004: http://www.google.com/search?hl=en&lr=&ie=UTF8&oi=defmore&q=define:-News+group

67. What is a Search Engine? Metamend.inc, 2004, visited at: 12/6/2004: http://www.metamend.com/search-engine-map.html

68. Moore, Sharon. Image Retrieval, [w. p.], 2001, visited at: 10/10/2003: http://tlc.dcs.gla.ac.uk/students/level3/imagedb/node13.html

رابعا: مواقع الويب

69. 911digitalarchive.http://911digitalarchive.org/
70. Accuweather.http://accuweather.ap.org/
71. Afp.http://www.afp.com/arabic/products/?pid=image
72. Ajkids.http://www.ajkids.com
73. Allanimation.http://www.allanimation.com/
74. Alltheweb.http://www.alltheweb.com
75. Almashriq.http://almashriq.hiof.no
76. Altavista.http://www.altavista.com
77. Altavista.http://www.altavista.com/mutimedia/default?qbmode=
78. Animalpicturesarchive.http://www.animalpicturesarchive.com/
79. Animationconnection.http://www.animationconnection.com/
80. Animationusa.http://www.animationusa.com/
81. Arlisna.http://www.arlisna.org/
82. Astronomy.http://www.astronomy.ca/images/
83. Business.http://www.business.com/
84. Canadaspace.http://www.canadaspace.com
85. Cartoonbank.http://www.cartoonbank.com /
86. Catalog.lapl.http://catalog.lapl.org/a_photo.html
87. Corbis.www.corbis.com
88. Ditto.www.ditto.com
89. Egyptsearch.http://www.egyptsearch.com/search/
90. Eserver.http://www.eserver.org/
91. Findlaw.http://findlaw.com
92. Fotosearch.HTTP://fotosearch.com
93. Fotosearch.www.fotosearch.com
94. Fotosearch.www.fotosearch.com
95. Freefoto.www.freefoto.com
96. Free-graphics.http://www.free-graphics.com/
97. Freeimages.http://www.freeimages.co.uk
98. Froogle.http://www.froogle.com
99. Google.www.google.com
100. Iconbank.http://iconbank.com/gx/index.htm
101. Infotoday.http://www.infotoday.com/searcher/may02/mattison.htm
102. Invisibleweb.http://www.Invisibleweb.com
103. Jewishhistory.huji.http://jewishhistory.huji.ac.il/links/maps.htm

104. KidsClick.http://www.kidsClick.com
105. Lcweb2.loc.http://lcweb2.loc.gov/ammem/gmdhtml/gmdhome.html
106. Lib.utexas.http://www.lib.utexas.edu/Libs/PCL/Map_collection/historical/history_-main.html
107. Library.uthscsa.http://www.library.uthscsa.edu/
108. Lifeart.http://www.lifeart.com
109. Luxcentral.http://www.luxcentral.com/stamps/LuxStampsDB.htm
110. Mamma.http://www.mamma.com
111. Medicine.ucsd.http://medicine.ucsd.edu/clinicalimg/
112. MetaCrawler.http://www.metaCrawler.com
113. Metrokc.http://www.metrokc.gov/gis/
114. Midisearch.http://midisearch.net/
115. Mltavista.http://www.mltavista.com
116. Msnshopping.http://www.msnshopping.com
117. Multimedia.alltheweb.http://multimedia.alltheweb.com
118. Multimedia.lycos.http://multimedia.lycos.com/
119. Music-finder.http://www.music-finder.net/
120. Mwsearch.http://www.mwsearch.com/
121. Naturepl.http://www.naturepl.com/frmsns.html
122. News.altavista.http://news.altavista.com/
123. News.google.http://news.google.com/
124. News.yahoo.http://news.yahoo.com/
125. Nmmagazine.http://www.nmmagazine.com/cgi-bin/links.cgi
126. Open.http://www.open.org/~library/fotofind.html
127. Photoartofnature.http://www.photoartofnature.com/index.html
128. Photostore.http://www.photostore.com/psx/index.html
129. Picsearch.www.picsearch.om
130. Pratique.http://www.pratique.fr/
131. Rock-n-reel.http://www.rock-n-reel.com/
132. Rps.http://www.rps.org
133. Seaborg.nmu.http://seaborg.nmu.edu/earth/Life.html
134. Search.singingfish.http://search.singingfish.com/sfw/home.jsp
135. Search4science.http://www.search4science.com/
136. Searchdesk.http://www.searchdesk.com/
137. Searcheurope.http://www.searcheurope.com
138. Sportec.http://www.sportec.com/
139. Tdfilm.http://www.tdfilm.com/

140. Torro.http://www.torro.org.uk/photos.htm
141. Usda.http://www.usda.gov/oc/photo/agri11.htm
142. Visibleearth.http://www.visibleearth.nasa.gov/
143. Visoo.http://www.visoo.com
144. Web.onramp. http://web.onramp.ca/phsc/]
145. Webdata.http://www.Webdata.com
146. Webmasters.http://webmasters.abondance.com/
147. Webmovie.http://www.webmovie.com/
148. Wordweb.http://wordweb.com/docs/gallery/geo/
149. Worldbank.http://www.worldbank.org/html/extdr/emps/photo.htm
150. Yahoo.http://www.yahoo.com
151. Yahoo.www.yahoo.com
152. Yahooligans.http://www.Yahooligans.com
153. Yahooshopping.http://www.yahooshopping.com
154. Yupimsn.http://www.yupimsn.com/index.asp



الملاحـــــق

الملحق رقم (١)
قائمة مصطلحات الدراسة
(إنجليزي - عربي)

المصطلح العربي	المصطلح الإنجليزي
فراغ العنوان	Address Bar
الرسوم الحيوية	Animation
التكشيف بالتعيين	Assigned Indexing Terms
بطاقة الوصف البيليوجرافي	Bibliographic Description Card
المنطق البوليني	Boolean Logic
معاملات المنطق البوليني	Boolean Operators
البحث بحساسية حالة الحروف	Case Ssensitive
البحث المبني على المحتوى	Concept-Based Searching
استرجاع الصور المبني على المحتوى	Content Based Image Retrieval
حقوق نشر الصورة	Copy Rights
تاريخ التحميل على الخط المباشر	Date Placed online
تاريخ إنشاء التسجيلة	Date Record Created
الويب العميقة	Deep Web
عمق عمل الزاحف	Depth of Crawl
استرجاع الصور بتصفح المصطلحات	Direct Text Image Retrieval
روابط غير نشطة	Disconnected Links
اسم نطاق الموقع	Domain Name
البحث بحقول الوصف	Field Searching
شكل ملف الصورة	File Shape
تصفح ملفات الصور	Folder Viewers
مدى تكرار عملية الزحف	Frequency of Crawl
نظم تصفح الصور الكاملة	Full-Image-Management Systems
الصفحة الدليلية	Home page
بروتوكول نقل النص الفائق	Hyper Text Transfer Protocol (http)

Hypertext Markup Language (html)	لغة ترميز النص الفائق
Image Content Analysis	تحليل محتوى الصور
Image Management Systems	نظم إدارة الصور
Image Media	وسيط الصورة
Inside Pictorial Image Retrieval	نظم استرجاع الصور الداخلية
Invisible Web	الويب الخفية
Meta Search Engines	محركات البحث المتعددة
Meta Tags	حقول صفحات الويب
Metadata	الميتاداتا
Mirror Sites	الصفحات الاحتياطية
Moving Visual Materials	المواد المرئية المتحركة
Outside Pictorial Image	نظم تصفح الصور الخارجية
Photo Format	نوع ملف الصورة
Phrase Searching	البحث بالعبارات
Pictorial Image Retrieval	نظم التصفح المصورة
Proximity Searching	البحث بتقارب الكلمات
Sketch Image Retrieval	استرجاع الصور بالنماذج التصويرية
Still Visual Materials	المواد المرئية الثابتة
Surrounding Text	النص المحيط بالصور
Tagging	ترميز الملفات الإلكترونية
Text Based Image Retrieval	استرجاع الصور المبني على النص
Text Media	وسيط النص
Textual Image Browsing	تصفح الصور بالمصطلحات
Texture	بنية ملفات الصور
The Opaque Web	الويب غير الشفافة
The Private Web	الويب الخاصة
The Proprietary Web	الويب ذات الملكية
Thumbnail	اللقطات
Union Resource Locator (URL)	مصدر المعلومات الموحد
Visual Browsing	التصفح المرئي
Web Browsers	متصفحات الويب
Web page	صفحة الويب
Wild card	البحث الحر

الملحق (٢)

أولا نماذج الوصف الفني للصور الرقمية

Top Level Metadata Description for Complete Image

Title: A selector and his family, probably in the Beenleigh district, 1872

Creator: William Boag

Subject: Photograph collection - Queensland

Description: The difficulties faced by a family in the Queensland bush included poor roads, an unreliable mail service and dense, vine-matted scrub. For many years, a selector's staple diet was salted meat (salt horse) and pumpkins. For several months, a woman and her children might be alone in their stringy-bark hut while her husband went off to split shingles or to earn extra money on a cattle property.

Date.created: 1872

Date.recordCreated: 1996

Date.placedOnline: 1997

Publisher: State Library of Queensland

Type: image.photograph

Format: image/jpg

Format.fileSize: 50.6Kb

Format.dimensions: 672 x 512

Format.colorpalette: grayscale

Identifier: http://archive.dstc.edu.au/RDU/SLQ/boag/20248.jpg

Source: BOAG negative no. 906

Language: en

Relation.isPartOf: http://www.slq.qld.gov.au/jol/boag.htm

Relation.hasParts: Region1

Relation.hasFormat: http://archive.dstc.edu.au/RDU/SLQ/boag/20248.gif

Coverage: Beenleigh region, Queensland, 1872

Rights: http://www.slq.qld.gov.au/cright.htm

Secondary Level Metadata for Region1

Identifier: Region1

Title: Annie Dickson

Description: Wife of James Dickson and mother to their 13 children.

Coverage.rect: 495,207,546,263

Relation.isPartOf: http://archive.dstc.edu.au/RDU/SLQ/boag/20248.j



الملحق (٣)
نماذج الوصف الفني للصور التقليدية

CREATOR

Keenan's Gallery, 118 West Fourth St., Cincinnati, O.

TITLE

[Three-quarter portrait of Adjutant William Blackford Stephenson, 1840-1879]

DESCRIPTION

1 photoprint : albumen, 10 x 6 cm. (Carte de visite)

SUBJECTS

Military officers – Ohio

Ohio Infantry, 53rd Regiment

COLLECTION

Higby Family Photograph Collection

LOCATION

P20/1/6

COPYRIGHTS

Copyrights have not been dedicated to the public. Consideration of the requirements of copyrights is the responsibility of the author and publisher.

NOTES

verso: "Camp Denison/March 7th, 1864/Adj Stephenson/53rd Reg com"

CONTROL NO.

CWI01.txt



CREATOR

unknown

TITLE

[Group portrait of 20 officers of the 82nd O.V.I., posing on steps of the Ohio State House]

DESCRIPTION

Silver-gelatin print, 12 x 16 cm.

SUBJECTS

Ohio infantry, 82nd Regiment

Military officers - Ohio

COLLECTION

F. & R. Lazarus Photographic Collection

LOCATION

P92/11/#213

COPYRIGHTS

Copyrights have not been dedicated to the public. Consideration of the requirements of copyrights is the responsibility of the author and publisher.

NOTES

verso: "82nd O.V.I."

CONTROL NO.

CWI02.txt



CREATOR

Gurney & Son, Fifth Ave., N.Y.

TITLE

[Portrait of General George H. Thomas]

DESCRIPTION

Cabinet photograph ; 17 x 11 cm.

SUBJECTS

Military officers - Ohio - Generals - Military service

Thomas, George Henry, 1816-1870

COLLECTION

John H. Klippart Family Collection

LOCATION

P157/1/3

COPYRIGHTS

Copyrights have not been dedicated to the public. Consideration of the requirements of copyrights is the responsibility of the author and publisher.

NOTES

verso: "Gen. Thomas"

CONTROL NO.

CWI03.txt

الملحق (٤)

هيئات الصور الرقمية

هيئة I3A

يأتي هذا الاختصار للتعبير عن The International Imaging Industry Association، وهي مؤسسة تعمل على دعم ملفات الصور الرقمية، بشكل غير هادف للربح.وتتكون هذه الهيئة من تكامل مجموعة كبيرة من الشركات حول العالم تعمل على وضع المعايير و مواصفات التعامل مع الصور. ويمكن القول أيضًا إن هذه الهيئة هي نتاج اندماج مجموعة من الهيئات مثل Digital Imaging Group (DIG) and the Photographic and Imaging Manufacturers Association (PIMA). وقد تواصل عملها لمدة ٦٠ عامًا في هذا المجال. ثم حملت هذه الهيئة على عاتقها منذ ذلك التعامل مع المواصفات الملائمة لبناء صناعة التصوير والصور الرقمية.

هيئة NISO

يعبر هذا الاختصار عن National Information Standards Organization،وقامت هذه الهيئة اعتمادًا على تفويض هيئة American National Standards Institute (ANSI). تعمل هذه المؤسسة على إصدار معايير يمكنها التطبيق على البيئتين التقليدية والرقمية يمكن معها تلبية احتياجات المعلومات المختلفة ؛ بما في ذلك من عمليات الاسترجاع والحفظ والتخزين والميتاداتا.وقد بدأت هذه الهيئة منذ عام ١٩٣٩م ثم تغيرت إلى هيئة غير هادفة للربح في عام ١٩٨٣ تحمل الطابع التعليمي، ثم استقرت على اسمها الحالي) (NISO) بعد ذلك بعام واحد.

هيئة VRA

ينطوي هذا الاختصار على مفهوم Visual Resources Association وقد بدأ الإعداد لهذه الهيئة منذ انعقاد اللقاء السنوي College Art Association conferences (CAA) في

عام ١٩٦٨م، ثم خرجت توصية لأعضاء اللقاء السنوي لهيئة Art Libraries Society of North America (ARLIS/NA). المنعقد في عام ١٩٧٠م على أن يتم البدء في تشكيل لجنة أو هيئة من الأعضاء للتعامل مع موضوعات إدارة وتنظيم المعلومات المصورة. وفي نهاية عقد السبعينات وبداية عام ١٩٨٠م بدأت هيئتا (CAA)، (ARLIS/NA) العمل على وضع المراحل الأولى لإقامة هذه الهيئة معتمدة على مجموع ٦٠٠ عضو حول العالم. وعملت هذه الهيئة بعد ذلك على إصدار نشرة خاصة بالهيئة تتناول منشورات المؤتمر السنوي بالإضافة إلى الاهتمام بموضوعات تنظيم وإدارة المواد المصورة مركزة على مواد الشرائح و الصور الفوتوغرافية.

هيئة TASI

تعد هذه الهيئة أهم هيئات التعامل مع الصور الرقمية على الويب، ويشير اختصار الهيئة إلى مفهوم The Technical Advisory Service for Images وقد قامت هذه الهيئة معتمدة على دعم هيئة JISC Joint Information Systems Committee. وتقدم هذه الهيئة خدماتها وإرشاداتها في مجال الصور الرقمية، فضلا عن الخدمات التعليمية في هذا المجال مع التركيز على بعض الموضوعات منها :

- ✓ إنتاج الصور الرقمية.
- ✓ توصيل الصور الرقمية إلى مستخدميها.
- ✓ تطبيقات الصور الرقمية في مجالات التعليم والتجارة والبحث العلمي.
- ✓ إدارة الكثير من المشروعات الخاصة بالتصوير والصور الرقمية.

وتعمل هذه الهيئة على تحقيق مجموعة من الأهداف مثل :

- ☒ دعم وتشجيع إنتاج واستخدام الصور الرقمية عالية الكثافة في مجالات البحث والتعليم.
- ☒ ترويج العمل باستخدام الصور الرقمية عن طريق إكسابها المواصفات الملائمة لذلك.
- ☒ دعم كل ما يتعلق بالصور الرقمية سواء أكان الإرشاد أم الاستشارات أم التدريب أم المساندة.

هيئة Visual Arts Data Service

تتولى هذه الهيئة مهمة البحث العلمي الخاص بالمواد المصورة الرقمية، فضلا عن دعم إنتاج هذه المواد ودعم خصائص الحفظ طويل المدى لها. وتعمل هذه الهيئة على تحقيق مجموعة من الأهداف منها ما يلي:-

- دعم إنتاج وتجهيز المواد المصورة الرقمية للبحث والاستخدام على شبكة الويب.
- دعم حفظ المواد الرقمية المصورة للتأكيد على المدى الممتد لاستخدامها.
- العمل على ترويج وممارسة التعامل مع المواد المصورة الرقمية.

وتقدم هذه الهيئة أيضًا مجموعة من الخدمات مثل :

- ☒ خدمة الوصول إلى المواد المصورة الرقمية.
- ☒ حفظ مجموعات المواد المصورة الرقمية لفترة تتيح استخدام ممتد لهذه المواد.
- ☒ تقدم خدمة التواصل العلمي والمشروعات العلمية في قطاعات دراسة المواد المصورة الرقمية.



الملحق (٥)

واجهات محركات بحث عينة الدراسة

يشمل هذا الملحق على الترتيب ما يلي :-

١. البحث بآليات المنطق البوليني في محركات Altavista, Google.

٢. آلية البحث بحساسية الحروف لمحرك البحث Altavista.

٣. البحث بآليات اللغة العربية داخل محركات البحث Google, Alltheweb.



altavista Web **Images** MP3/Audio Video News

Egypt not pyramids [Find]

Find: ☑ Photos ☑ Graphics ☐ Buttons/Banners

Color: All Colors **Sources:** All Sources **Sizes:** All Sources

















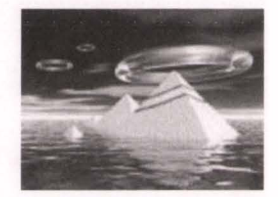
altavista Web **Images** MP3/Audio Video News

Egypt not pyramids [Find]

Find: ☑ Photos ☑ Graphics ☐ Buttons/Banners

Color: All Colors **Sources:** All Sources **Sizes:** All Sources

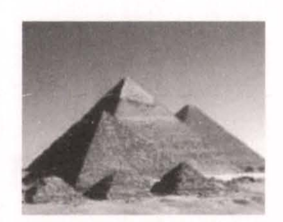


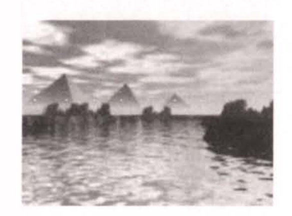
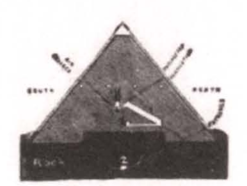










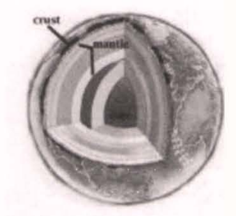

altavista Web **Images** MP3/Audio Video News

EARTH **Find**

Find: ☑Photos ☑Graphics ☐Buttons/Banners

Color: All Colors **Sources:** All Sources **Sizes:** All Sources



















advanced search :: customize preferences :: submit site :: help

القمر

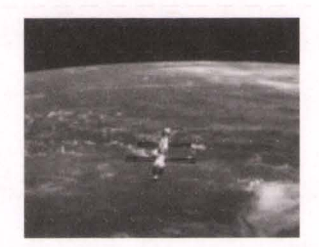
SEARCH

| Web | News | **Pictures** | Video | Audio |

1 - 12 of 8,238 Results for القمر

Offensive content filter: On - Off





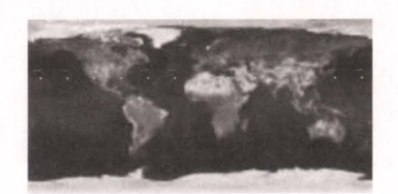












ويب **صور** مجموعات الدليل

Google
بحث عن الصور

سيارة [ابحث] البحث المتقدم عن الصور
التفضيلات

صور النتائج ٢١-٤٠ من حوالي ٢،٢٥٠ لـ طائرة (الوقت المستغرق ٠،١٨)

عرض: **كل الأحجام** - كبير - متوسط - صغير















الكتاب:

- شكلت الصور الرقمية في بيئة الويب واحدة من أهم وسائط حمل المعلومات الرقمية. وتنوعت المعلومات المرئية داخل الصور عامة والثابتة خاصة بين المواد الإعلامية والصور العلمية والصور الفنية العامة. وقد أخذت أعداد ملفات الصور الرقمية في تزايد مستمر إلى أن شكلت كماً هائلاً من مصادر المعلومات الرقمية المرئية التي تحتاج إلى معالجة فنية وتنظيم واسترجاع ذات طبيعة خاصة. غير أن الويب لم تقدم المعايير الملائمة للمعالجة الفنية أو التنظيم للصور الرقمية، كما أن أدوات تنظيم وبحث الصور الرقمية (محركات بحث الصور) لم تستطع تقديم الآليات الملائمة للتعامل مع وسيط الصورة بوصفه مصدر معلومات رقمياً، سواء كان ذلك في المعالجة والتنظيم أم البحث والاستدعاء لملفات الصور الرقمية.
- عمل هذا البحث على دراسة خصائص بيئة الويب باعتبارها بيئة لمعالجة واسترجاع الصور الرقمية الثابتة، إلى جانب دراسة خصائص الصورة الرقمية الثابتة بوصفها وسيطاً لحمل المعلومات المرئية؛ من حيث عناصرها الخارجية الشكلية أو الداخلية الموضوعية، بالإضافة إلى دراسة آليات عمل محركات بحث الصور الثابتة وخصائص المعالجة الفنية والتنظيم داخلها.
- كما اهتم البحث بصياغة مواصفات محركات بحث الصور الثابتة بما يحاكي ويلائم خصائص اللغة العربية باعتبارها إحدى لغات البحث والاسترجاع للمصادر الرقمية في بيئة الويب.
- انتهى البحث إلى وضع نموذج لملامح عمل محركات بحث الصور الرقمية بما يدعم المعالجة والتنظيم والاستدعاء وفقاً لطبيعة اللغة العربية.

المؤلف:

سيد ربيع سيد إبراهيم

- ماجستير في علم المكتبات والمعلومات عام ٢٠٠٥م.
- اجتاز دورات علمية عدة.
- له أبحاث علمية منشورة في مجال التخصص.
- يعمل الآن مدرساً مساعداً بكلية الآداب جامعة بني سويف ـ قسم المكتبات والوثائق.

ردمك: ٨-٢٨٨-٠٠-٩٩٦٠

ISBN: 9960 - 00 - 288 - 8